PCT/AU03/01166

PRIORITY DOCUMENT

SUBMITTED OR TRANSMITTED IN COMPLIANCE WITH RULE 17.1(a) OR (b)

Patent Office
Canberra

I, JONNE YABSLEY, TEAM LEADER EXAMINATION SUPPORT AND SALES hereby certify that annexed is a true copy of the Provisional specification in connection with Application No. 2002951346 for a patent by GARVAN INSTITUTE OF MEDICAL RESEARCH as filed on 05 September 2002.

WITNESS my hand this
Fifteenth day of September 2003

JONNE YABSLEY
TEAM LEADER EXAMINATION
SUPPORT AND SALES

AUSTRALIA

Patents Act 1990

Garvan Institute of Medical Research

PROVISIONAL SPECIFICATION

Invention Title:

Diagnosis of Ovarian Cancer

The invention is described in the following statement:

METHODS OF DIAGNOSIS OF OVARIAN CANCER, COMPOSITIONS AND METHODS OF SCREENING FOR MODULATORS OF OVARIAN CANCER

Field of the invention

The present invention relates to the identification of nucleic acid and protein expression profiles and nucleic acids, products, and antibodies thereto that are involved in ovarian cancer; and to the use of such expression profiles and compositions in the diagnosis, prognosis and therapy of ovarian cancer. The invention further relates to methods for identifying and using agents and/or targets that inhibit ovarian cancer.

Background of the invention

Ovarian cancer is the fourth most frequent cause of cancer death in females and in the United States, and accounts for approximately 13,000 deaths annually. Furthermore, ovarian cancer remains the number one killer of women with gynaecological malignant hyperplasia and the incidence is rising in industrialized countries. The etiology of the neoplastic transformation remains unknown although there is epidemiological evidence for an association with disordered endocrine function. The incidence of ovarian carcinoma is higher in nulliparous females and in those with early menopause.

Approximately 75% of women diagnosed with such cancers are already at an advanced stage (III and IV) of the disease at their initial diagnosis. During the past 20 years, neither diagnosis nor five year survival rates have greatly improved for these patients. This is substantially due to the high percentage of high-stage initial detection of the disease. There is therefore a need to develop new markers that improve early diagnosis and thereby reduce the percentage of high-stage initial diagnoses.

Summary of the invention

The present invention is based on the identification of a number of genes whose expression is altered in individuals with ovarian cancer versus individuals who do not have ovarian cancer. The particular genes are identified in the Tables.

Accordingly, the present invention provides in a first aspect, a method of detecting an ovarian cancer-associated transcript in a biological sample, the method comprising contacting the biological sample with a polynucleotide that

selectively hybridises to a sequence at least 80% identical to a sequence as shown in the Tables.

Preferably the polynucleotide selectively hybridises to a sequence at least 95% identical to a sequence as shown in the Tables. More preferably, the polynucleotide comprises a sequence as shown in the Tables.

In one embodiment, the method further comprises the step of amplifying nucleic acids before the step of contacting the biological sample with the polynucleotide.

In a preferred embodiment, the polynucleotide is immobilised on a solid surface.

In a related aspect, the present invention provides a method of detecting an ovarian cancer-associated polypeptide in a biological sample the method comprising contacting the biological sample with an antibody which binds specifically to the ovarian cancer-associated polypeptide in the biological sample, the polypeptide being encoded by a polynucleotide that selectively hybridises to a sequence at least 80% identical to a sequence as shown in the Tables.

In a further related aspect, the present invention provides a method of detecting an ovarian cancer-associated antibody in a biological sample the method comprising contacting the biological sample with a polypeptide encoded by a polynucleotide that selectively hybridises to a sequence at least 80% identical to a sequence as shown in the Tables, wherein the polypeptide specifically binds to the ovarian cancer-associated antibody.

Preferably, in the above methods, the biological sample is contacted with a plurality of the polynucleotides, polypeptides or antibodies referred to above.

The biological sample is typically a tissue sample or a blood sample.

In one embodiment, the biological sample is from a patient undergoing a therapeutic regimen to treat ovarian cancer. In an alternative embodiment, the biological sample is from a patient suspected of having ovarian cancer.

In a second aspect, the present invention a method of monitoring the efficacy of a therapeutic treatment of ovarian cancer, the method comprising the steps of:

(i) providing a biological sample from a patient undergoing the therapeutic treatment; and

(ii) determining the level of a ovarian cancer-associated transcript in the biological sample by contacting the biological sample with a polynucleotide that selectively hybridises to a sequence at least 80% identical to a sequence as shown in the Tables, thereby monitoring the efficacy of the therapy.

Preferably the method further comprises the step of: (iii) comparing the level of the ovarian cancer-associated transcript to a level of the ovarian cancer-associated transcript in a biological sample from the patient prior to, or earlier in, the therapeutic treatment.

In a related aspect, the present invention provides a method of monitoring the efficacy of a therapeutic treatment of ovarian cancer, the method comprising the steps of:

(i) providing a biological sample from a patient undergoing the therapeutic treatment; and

(ii) determining the level of a ovarian cancer-associated antibody in the biological sample by contacting the biological sample with a polypeptide encoded by a polynucleotide that selectively hybridises to a sequence at least 80% identical to a sequence as shown in the Tables, wherein the polypeptide specifically binds to the ovarian cancer-associated antibody, thereby monitoring the efficacy of the therapy.

Preferably the method further comprises the step of: (iii) comparing the level of the ovarian cancer-associated antibody to a level of the ovarian cancer-associated antibody in a biological sample from the patient prior to, or earlier in, the therapeutic treatment.

In a further related aspect, the present invention provides a method of monitoring the efficacy of a therapeutic treatment of ovarian cancer, the method comprising the steps of:

(i) providing a biological sample from a patient undergoing the therapeutic treatment; and

(ii) determining the level of a ovarian cancer-associated polypeptide in the biological sample by contacting the biological sample with an antibody, wherein the antibody specifically binds to a polypeptide encoded by a polynucleotide that selectively hybridises to a sequence at least 80% identical to a sequence as shown in the Tables, thereby monitoring the efficacy of the therapy.

Preferably the method further comprises the step of: (iii) comparing the level of the ovarian cancer-associated polypeptide to a level of the ovarian cancer-associated polypeptide in a biological sample from the patient prior to, or earlier in, the therapeutic treatment.

Preferably, in the above methods, the biological sample is contacted with a plurality of the polynucleotides, polypeptides or antibodies referred to above.

In a third aspect, the present invention provides an isolated polynucleotide selected from the group consisting of:

(a) polynucleotides comprising a nucleotide sequence as shown in the Tables, or the complement thereof;

(b) polynucleotides comprising a nucleotide sequence capable of selectively hybridising to a nucleotide sequence as shown in the Tables;

(c) polynucleotides comprising a nucleotide sequence capable of selectively hybridising to the complement of a nucleotide sequence as shown in the Tables; and

(d) polynucleotides comprising a polynucleotide sequence which is degenerate as a result of the genetic code to the polynucleotides defined in (a), (b) or (c).

Preferred polynucleotides comprise a polynucleotide sequence as shown in the Tables or a sequence having at least 80% homology thereto.

The present invention also provides a nucleic acid vector comprising a polynucleotide of the invention. In one embodiment, the polynucleotide is operably linked to a regulatory control sequence capable of directing expression of the polynucleotide in a host cell.

The present invention further provides a host cell comprising a vector of the invention.

In a fourth aspect, the present invention provides an isolated polypeptide which is encoded by a polynucleotide of the invention. The present invention also provides an isolated polypeptide encoded by a polynucleotide that selectively hybridises to a sequence at least 80% identical to a sequence as shown in the Tables.

In a fifth aspect the present invention provides an antibody that binds specifically a polypeptide of the invention.

In a sixth aspect, the present invention provides a method for identifying a compound that modulates an ovarian cancer-associated polypeptide, the method comprising the steps of:

(i) contacting the compound with a ovarian cancer-associated polypeptide, the polypeptide encoded by a polynucleotide that selectively hybridises to a sequence at least 80% identical to a sequence as shown in the Tables; and

(ii) determining the functional effect of the compound upon the polypeptide.

The functional effect may, for example, be a physical effect or a chemical effect.

In one embodiment, the functional effect is determined by measuring ligand binding to the polypeptide.

In a particular embodiment, the polypeptide is expressed in a eukaryotic host cell or cell membrane.

Preferably the polypeptide is recombinant.

In another aspect, the present invention provides a method of inhibiting proliferation of a ovarian tumour cell, which method comprises contacting said cell with a compound identified using the method of the sixth aspect of the invention.

In a further aspect, the present invention provides a method of inhibiting proliferation of a ovarian cancer-associated cell to treat ovarian cancer in a patient, the method comprising the step of administering to the patient a therapeutically effective amount of a compound identified using the method of the sixth aspect of the invention.

In a seventh aspect, the present invention provides a drug screening assay comprising the steps of:

(i) administering a test compound to a mammal having ovarian cancer or a cell isolated therefrom;

(ii) comparing the level of gene expression of a polynucleotide that selectively hybridises to a sequence at least 80% identical to a sequence as shown in the Tables in a treated cell or mammal with the level of gene expression of the polynucleotide in a control cell or mammal, wherein a test compound that modulates the level of expression of the polynucleotide is a candidate for the treatment of ovarian cancer.

Typically, the control is a mammal with ovarian cancer or a cell therefrom that has not been treated with the test compound. Alternatively, the control is a normal cell or mammal.

The present invention also provides a method for treating a mammal having ovarian cancer comprising administering a compound identified by the assay of the seventh aspect of the invention.

In an eighth aspect, the present invention provides a pharmaceutical composition for use in treating a mammal having ovarian cancer, the composition comprising a compound identified by the method of the sixth aspect of the invention, or the assay of the seventh aspect of the invention, and a physiologically acceptable carrier or diluent.

In a ninth aspect, the present invention provides an assay device comprising a plurality of different polynucleotides of the invention to a solid phase. Preferably, the solid phase is a substantially planar chip.

In a related aspect, the present invention provides an assay device comprising a plurality of different antibodies of the invention immobilised to a solid phase. Preferably, the solid phase is a substantially planar chip.

The present invention also provides the use of a device of the invention in a method of the first or second aspect of the invention or an assay method of the seventh aspect of the invention.

In a tenth aspect, the present invention provides a non-human transgenic animal which is transgenic by virtue of comprising a polynucleotide of the invention as well as transgenic "knock-out" animals that have a disrupted gene corresponding to a polynucleotide of the invention.

Preferred sequences in relation to any or all of the aspects of the invention referred to above, and throughout the specification, are those that are listed in one or more of the Tables as having a P value of less than 0.05, more preferably a P value of less than 0.01. Similarly, preferred sequences are those referred to in Table 3B as having a ratio of expression between ovarian patients and normal patients of at least 5, more preferably at least 6, 7, 8, 9 or 10, and those sequences referred to in Table 3B as having a ratio of expression between ovarian patients and normal patients of less than –5 or –10.

Throughout this specification, preferred aspects and embodiments apply, as appropriate, separately, or in combination, to other aspects and embodiments, *mutatis mutandis*, whether or not explicitly stated as such.

Detailed description of the invention

Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art (e.g. in cell culture, molecular genetics, nucleic acid chemistry, hybridization techniques and biochemistry). Standard techniques are used for molecular, genetic and biochemical methods (see generally, Sambrook *et al.*, Molecular Cloning: A Laboratory Manual, 3rd ed. (2001) Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. and Ausubel *et al.*, Short Protocols in Molecular Biology (1999) 4th Ed, John Wiley & Sons, Inc. - and the full version entitled Current Protocols in Molecular Biology, which are incorporated herein by reference) and chemical methods.

Throughout this specification the word "comprise", or variations such as "comprises" or "comprising", will be understood to imply the inclusion of a stated element, integer or step, or group of elements, integers or steps, but not the exclusion of any other element, integer or step, or group of elements, integers or steps.

The present invention provides nucleic acid and protein sequences that are differentially expressed in ovarian cancer, herein termed "ovarian cancer sequences." As outlined below, ovarian cancer sequences include those that are up-regulated (i.e., expressed at a higher level) in ovarian cancer, as well as those that are down-regulated (i.e., expressed at a lower level). In a preferred embodiment, the ovarian cancer sequences are from humans; however, as will be appreciated by those in the art, ovarian cancer sequences from other organisms may be useful in animal models of disease and drug evaluation; thus, other ovarian cancer sequences are provided, from vertebrates, including mammals, including rodents (rats, mice, hamsters, guinea pigs, etc.), primates, farm animals (including sheep, goats, pigs, cows, horses, etc.) and pets, e.g., (dogs, cats, etc.).

Ovarian cancer sequences from other organisms may be obtained using the techniques outlined below. Ovarian cancer sequences can include both nucleic acid and amino acid sequences. As will be appreciated by those in the art and is more fully outlined below, ovarian cancer nucleic acid sequences are useful in a variety of applications, including diagnostic applications, which will detect naturally occurring nucleic acids, as well as screening applications; e.g., biochips comprising nucleic acid probes or PCR microtitre plates with selected probes to the ovarian cancer sequences can be generated.

An ovarian cancer sequence can be initially identified by substantial nucleic acid and/or amino acid sequence homology to the ovarian cancer sequences outlined herein. Such homology can be based upon the overall nucleic acid. or amino acid sequence, and is generally determined as outlined below, using either homology programs or hybridization conditions.

For identifying ovarian cancer-associated sequences, the ovarian cancer screen typically includes comparing genes identified in different tissues, e.g., normal and cancerous tissues, or tumour tissue samples from patients who have metastatic disease vs. non metastatic tissue. Other suitable tissue comparisons include comparing ovarian cancer samples with metastatic cancer samples from other cancers, such as lung, breast, gastrointestinal cancers, ovarian, etc. Samples of different stages of ovarian cancer, e.g., survivor tissue, drug resistant states, and tissue undergoing metastasis, are applied to biochips comprising

nucleic acid probes. The samples are first microdissected, if applicable, and treated as is known in the art for the preparation of mRNA. Suitable biochips are commercially available, e.g. from Affymetrix. Gene expression profiles as described herein are generated and the data analyzed.

In one embodiment, the genes showing changes in expression as between normal and disease states are compared to genes expressed in other normal tissues, preferably normal ovarian, but also including, and not limited to lung, heart, brain, liver, breast, kidney, muscle, colon, small intestine, large intestine, spleen, bone and placenta. In a preferred embodiment, those genes identified during the ovarian cancer screen that are expressed in any significant amount in other tissues are removed from the profile, although in some embodiments, this is not necessary. That is, when screening for drugs, it is usually preferable that the target be disease specific, to minimise possible side effects.

In a preferred embodiment, ovarian cancer sequences are those that are up-regulated in ovarian cancer; that is, the expression of these genes is higher in the ovarian cancer tissue as compared to non-cancerous tissue (see Table 1). "Up-regulation" as used herein means at least about a two-fold change, preferably at least about a three fold change, with at least about five-fold or higher being preferred. All Unigene cluster identification numbers and accession numbers herein are for the GenBank sequence database and the sequences of the accession numbers are hereby expressly incorporated by reference. Sequences are also available in other databases, e.g., European Molecular Biology Laboratory (EMBL) and DNA Database of Japan (DDBJ).

In another preferred embodiment, ovarian cancer sequences are those that are down-regulated in ovarian cancer; that is, the expression of these genes is lower in ovarian cancer tissue as compared to non-cancerous tissue (see Table 2 for example). "Down-regulation" as used herein often means at least about a 1.5-fold change more preferably a two-fold change, preferably at least about a three fold change, with at least about five-fold or higher being most preferred.

Preferred sequences are those referred to in the Tables that are listed in one or more of the Tables as having a P value of less than 0.05, more preferably a P value of less than 0.01. Similarly, preferred sequences are those referred to in Table 3B as having a ratio of expression between ovarian patients and normal patients of at least 5, more preferably at least 6, 7, 8, 9 or 10, and those sequences referred to in Table 3B as having a ratio of expression between ovarian patients and normal patients of less than –5 or –10.

Polynucleotides

Polynucleotides of the invention comprise the nucleic acid sequences whose nucleotide sequences are provided by reference to the public database accession numbers given in the Tables (referred to herein as the nucleotide sequences shown in the Tables), and sequences homologues thereto as well as variants, derivatives and fragments thereof.

The nucleotide sequences referred to in the Tables and homologues thereof, typically encode polypeptides. It will be understood by a skilled person that numerous different polynucleotides can encode the same polypeptide as a result of the degeneracy of the genetic code. In addition, it is to be understood that skilled persons may, using routine techniques, make nucleotide substitutions that do not affect the polypeptide sequence encoded by the polynucleotides of the invention to reflect the codon usage of any particular host organism in which the polypeptides of the invention are to be expressed.

Polynucleotides of the invention may comprise DNA or RNA. They may be single-stranded or double-stranded. They may also be polynucleotides which include within them synthetic or modified nucleotides. A number of different types of modification to oligonucleotides are known in the art. These include methylphosphonate and phosphorothioate backbones, addition of acridine or polylysine chains at the 3' and/or 5' ends of the molecule. For the purposes of the present invention, it is to be understood that the polynucleotides described herein may be modified by any method available in the art. Such modifications may be carried out in order to enhance the *in vivo* activity or life span of polynucleotides of the invention.

The terms "variant" or "derivative" in relation to the nucleotide sequences of the present invention include any substitution of, variation of, modification of, replacement of, deletion of or addition of one (or more) nucleic acid from or to the sequence provided that the resultant nucleotide sequence codes for a polypeptide having biological activity, preferably having substantially the same activity as the polypeptide sequences presented in the sequence listings.

With respect to sequence homology, preferably there is at least 75%, more preferably at least 85%, more preferably at least 90% homology to a sequence shown in the Tables herein over a region of at least 20, preferably at least 25 or 30, for instance at least 40, 60, 100, 500, 1000 or more contiguous nucleotides. More preferably there is at least 95%, more preferably at least 98%, homology. In one embodiment, homologues are naturally occurring sequences, such as orthologues, tissue-specific isoforms and allelic variants.

Homology comparisons can be conducted by eye, or more usually, with the aid of readily available sequence comparison programs. These commercially available computer programs can calculate % homology between two or more sequences.

% homology may be calculated over contiguous sequences, i.e. one sequence is aligned with the other sequence and each nucleotide in one sequence directly compared with the corresponding nucleotide in the other sequence, one base at a time. This is called an "ungapped" alignment. Typically, such ungapped alignments are performed only over a relatively short number of bases (for example less than 50 contiguous nucleotides).

Although this is a very simple and consistent method, it fails to take into consideration that, for example, in an otherwise identical pair of sequences, one insertion or deletion will cause the following nucleotides to be put out of alignment, thus potentially resulting in a large reduction in % homology when a global alignment is performed. Consequently, most sequence comparison methods are designed to produce optimal alignments that take into consideration possible insertions and deletions without penalising unduly the overall homology score. This is achieved by inserting "gaps" in the sequence alignment to try to maximise local homology.

However, these more complex methods assign "gap penalties" to each gap that occurs in the alignment so that, for the same number of identical amino acids, a sequence alignment with as few gaps as possible - reflecting higher relatedness between the two compared sequences - will achieve a higher score than one with many gaps. "Affine gap costs" are typically used that charge a relatively high cost for the existence of a gap and a smaller penalty for each subsequent residue in the gap. This is the most commonly used gap scoring system. High gap penalties will of course produce optimised alignments with fewer gaps. Most alignment programs allow the gap penalties to be modified. However, it is preferred to use the default values when using such software for sequence comparisons.

Calculation of maximum % homology therefore firstly requires the production of an optimal alignment, taking into consideration gap penalties. A suitable computer program for carrying out such an alignment is the GCG Wisconsin Bestfit package (University of Wisconsin, U.S.A.; Devereux *et al.*, 1984, Nucleic Acids Research 12:387). The default scoring matrix has a match value of 10 for each identical nucleotide and -9 for each mismatch. The default gap creation penalty is -50 and the default gap extension penalty is -3 for each nucleotide.

Examples of other software than can perform sequence comparisons include, but are not limited to, the BLAST package (see Ausubel *et al.*, 1999 *ibid* –

Chapter 18), FASTA (Atschul *et al.,* 1990, J. Mol. Biol., 403-410) and the GENEWORKS suite of comparison tools. Both BLAST and FASTA are available for offline and online searching (see Ausubel *et al.*, 1999 *ibid*, pages 7-58 to 7-60). However it is preferred to use the GCG Bestfit program.

Once the software has produced an optimal alignment, it is possible to calculate % homology, preferably % sequence identity. The software typically does this as part of the sequence comparison and generates a numerical result.

A preferred sequence comparison program is the GCG Wisconsin Bestfit program described above.

The present invention also encompasses nucleotide sequences that are capable of hybridising selectively to the sequences presented herein, or any variant, fragment or derivative thereof, or to the complement of any of the above. Nucleotide sequences are preferably at least 15 nucleotides in length, more preferably at least 20, 30, 40 or 50 nucleotides in length.

The term "hybridization" as used herein shall include "the process by which a strand of nucleic acid joins with a complementary strand through base pairing" as well as the process of amplification as carried out in polymerase chain reaction technologies.

Polynucleotides of the invention capable of selectively hybridising to the nucleotide sequences presented herein, or to their complement, will be generally at least 70%, preferably at least 80 or 90% and more preferably at least 95% or 98% homologous to the corresponding nucleotide sequences referred to in the Tables over a region of at least 20, preferably at least 25 or 30, for instance at least 40, 60, 100, 500, 1000 or more contiguous nucleotides.

The term "selectively hybridizable" means that the polynucleotide used as a probe is used under conditions where a target polynucleotide of the invention is found to hybridize to the probe at a level significantly above background. The background hybridization may occur because of other polynucleotides present, for example, in the cDNA or genomic DNA library being screening. In this event, background implies a level of signal generated by interaction between the probe and a non-specific DNA member of the library which is less than 10 fold, preferably less than 100 fold as intense as the specific interaction observed with the target DNA. The intensity of interaction may be measured, for example, by radiolabelling the probe, e.g. with ^{32}P.

Hybridization conditions are based on the melting temperature (Tm) of the nucleic acid binding complex, as taught in Berger and Kimmel (1987, Guide to Molecular Cloning Techniques, Methods in Enzymology, Vol 152, Academic Press, San Diego CA), and confer a defined "stringency" as explained below.

Maximum stringency typically occurs at about Tm-5°C (5°C below the Tm of the probe); high stringency at about 5°C to 10°C below Tm; intermediate stringency at about 10°C to 20°C below Tm; and low stringency at about 20°C to 25°C below Tm. As will be understood by those of skill in the art, a maximum stringency hybridization can be used to identify or detect identical polynucleotide sequences while an intermediate (or low) stringency hybridization can be used to identify or detect similar or related polynucleotide sequences.

In a preferred aspect, the present invention covers nucleotide sequences that can hybridise to the nucleotide sequence of the present invention under stringent conditions (e.g. 65°C and 0.1xSSC {1xSSC = 0.15 M NaCl, 0.015 M Na_3 Citrate pH 7.0}).

Where the polynucleotide of the invention is double-stranded, both strands of the duplex, either individually or in combination, are encompassed by the present invention. Where the polynucleotide is single-stranded, it is to be understood that the complementary sequence of that polynucleotide is also included within the scope of the present invention.

Polynucleotides which are not 100% homologous to the sequences of the present invention but fall within the scope of the invention can be obtained in a number of ways. Other variants of the sequences described herein may be obtained for example by probing DNA libraries made from a range of individuals, for example individuals from different populations. In particular, given that that changes in the expression of nucleotide sequences of the invention correlate with ovarian cancer, characterisation of variant sequences in individuals suffering from ovarian cancer may be used to identify variations in the sequences of ovarian-cancer associated polynucleotides (and proteins) that predictive of and/or causative of, ovarian cancer.

Accordingly the present invention provides methods of identifying sequence variations that are associated with ovarian cancer which methods comprise determining all or part of the nucleotide sequence of a gene referred to in the Tables, derived from an individual suffering from ovarian cancer and comparing the sequence to that of the corresponding wild-type gene.

In addition, other viral/bacterial, or cellular homologues particularly cellular homologues found in mammalian cells (e.g. rat, mouse, bovine and primate cells), may be obtained and such homologues and fragments thereof in general will be capable of selectively hybridising to the sequences shown in the sequence listing herein. Such sequences may be obtained by probing cDNA libraries made from or genomic DNA libraries from other animal species, and probing such libraries with probes comprising all or part of the sequences referred to in the Tables under conditions of medium to high stringency. Similar considerations apply to obtaining

species homologues and allelic variants of the nucleotide sequences referred to in the Tables.

Variants and strain/species homologues may also be obtained using degenerate PCR which will use primers designed to target sequences within the variants and homologues encoding conserved amino acid sequences within the sequences of the present invention. Conserved sequences can be predicted, for example, by aligning the amino acid sequences from several variants/homologues. Sequence alignments can be performed using computer software known in the art. For example the GCG Wisconsin PileUp program is widely used.

The primers used in degenerate PCR will contain one or more degenerate positions and will be used at stringency conditions lower than those used for cloning sequences with single sequence primers against known sequences.

Alternatively, such polynucleotides may be obtained by site-directed mutagenesis of characterised sequences, such as the sequences referred to in the Tables. This may be useful where for example silent codon changes are required to sequences to optimise codon preferences for a particular host cell in which the polynucleotide sequences are being expressed. Other sequence changes may be desired in order to introduce restriction enzyme recognition sites, or to alter the property or function of the polypeptides encoded by the polynucleotides.

Polynucleotides of the invention may be used to produce a primer, e.g. a PCR primer, a primer for an alternative amplification reaction, a probe e.g. labelled with a revealing label by conventional means using radioactive or non-radioactive labels, or the polynucleotides may be cloned into vectors. Such primers, probes and other fragments will be at least 15, preferably at least 20, for example at least 25, 30 or 40 nucleotides in length, and are also encompassed by the term polynucleotides of the invention as used herein. Preferred fragments are less than 5000, 2000, 1000, 500 or 200 nucleotides in length.

Polynucleotides such as a DNA polynucleotides and probes according to the invention may be produced recombinantly, synthetically, or by any means available to those of skill in the art. They may also be cloned by standard techniques.

In general, primers will be produced by synthetic means, involving a step wise manufacture of the desired nucleic acid sequence one nucleotide at a time. Techniques for accomplishing this using automated techniques are readily available in the art.

Longer polynucleotides will generally be produced using recombinant means, for example using PCR (polymerase chain reaction) cloning techniques. This will involve making a pair of primers (e.g. of about 15 to 30 nucleotides) flanking a region of the sequence which it is desired to clone, bringing the primers into contact with mRNA or cDNA obtained from an animal or human cell, performing a polymerase chain reaction under conditions which bring about amplification of the desired region, isolating the amplified fragment (e.g. by purifying the reaction mixture on an agarose gel) and recovering the amplified DNA. The primers may be designed to contain suitable restriction enzyme recognition sites so that the amplified DNA can be cloned into a suitable cloning vector

Polynucleotides or primers of the invention may carry a revealing label. Suitable labels include radioisotopes such as ^{32}P or ^{35}S, enzyme labels, or other protein labels such as biotin. Such labels may be added to polynucleotides or primers of the invention and may be detected using by techniques known in the art.

Polynucleotides or primers of the invention or fragments thereof labeled or unlabeled may be used by a person skilled in the art in nucleic acid-based tests for detecting or sequencing polynucleotides of the invention in the human or animal body.

Such tests for detecting generally comprise bringing a biological sample containing DNA or RNA into contact with a probe comprising a polynucleotide or primer of the invention under hybridising conditions and detecting any duplex formed between the probe and nucleic acid in the sample. Such detection may be achieved using techniques such as PCR or by immobilising the probe on a solid support, removing nucleic acid in the sample which is not hybridised to the probe, and then detecting nucleic acid which has hybridised to the probe. Alternatively, the sample nucleic acid may be immobilised on a solid support, and the amount of probe bound to such a support can be detected. Suitable assay methods of this and other formats can be found in for example W089/03891 and W090/13667.

Tests for sequencing nucleotides of the invention include bringing a biological sample containing target DNA or RNA into contact with a probe comprising a polynucleotide or primer of the invention under hybridising conditions and determining the sequence by, for example the Sanger dideoxy chain termination method (see Sambrook et al.).

Such a method generally comprises elongating, in the presence of suitable reagents, the primer by synthesis of a strand complementary to the target DNA or RNA and selectively terminating the elongation reaction at one or more of an A, C, G or T/U residue; allowing strand elongation and termination reaction to occur;

separating out according to size the elongated products to determine the sequence of the nucleotides at which selective termination has occurred. Suitable reagents include a DNA polymerase enzyme, the deoxynucleotides dATP, dCTP, dGTP and dTTP, a buffer and ATP. Dideoxynucleotides are used for selective termination.

Tests for detecting or sequencing nucleotides of the invention in a biological sample may be used as part of the methods of the invention for detecting ovarian cancer-associated transcripts and monitoring the efficacy of treatment of patients suffering from ovarian cancer as will be described in more details below.

The probes/primers of the invention may conveniently be packaged in the form of a test kit in a suitable container. In such kits the probe may be bound to a solid support where the assay format for which the kit is designed requires such binding. The kit may also contain suitable reagents for treating the sample to be probed, hybridising the probe to nucleic acid in the sample, control reagents, instructions, and the like.

Preferably, a kit of the invention comprises primers/probes suitable for selectively detecting a plurality of sequences, more preferably for selectively detecting a plurality of sequences that are listed in one or more of the Tables as having a P value of less than 0.05, more preferably a P value of less than 0.01. Similarly, a kit of the invention preferably comprises primers suitable for selectively detecting a plurality of sequences referred to in Table 3B as having a ratio of expression between ovarian patients and normal patients of at least 5, more preferably at least 6, 7, 8, 9 or 10, and those sequences referred to in Table 3B as having a ratio of expression between ovarian patients and normal patients of less than –5 or –10.

DNA Array technologies

Techniques for producing immobilised arrays of DNA molecules have been described in the art. Generally, most prior art methods describe how to synthesise single-stranded nucleic acid molecule arrays, using for example masking techniques to build up various permutations of sequences at the various discrete positions on the solid substrate. U.S. Patent No. 5,837,832, the contents of which are incorporated herein by reference, describes an improved method for producing DNA arrays immobilised to silicon substrates based on very large scale integration technology. In particular, U.S. Patent No. 5,837,832 describes a strategy called "tiling" to synthesize specific sets of probes at spatially-defined locations on a substrate which may be used to produced the immobilised DNA arrays. U.S. Patent

No. 5,837,832 also provides references for earlier techniques that may also be used.

Thus DNA may be synthesised *in situ* on the surface of the substrate. However, DNA may also be printed directly onto the substrate using for example robotic devices equipped with either pins or piezo electric devices.

The plurality of polynucleotide sequences are typically immobilised onto or in discrete regions of a solid substrate. The substrate may be porous to allow immobilisation within the substrate or substantially non-porous, in which case the library sequences are typically immobilised on the surface of the substrate. The solid substrate may be made of any material to which polypeptides can bind, either directly or indirectly. Examples of suitable solid substrates include flat glass, silicon wafers, mica, ceramics and organic polymers such as plastics, including polystyrene and polymethacrylate. It may also be possible to use semi-permeable membranes such as nitrocellulose or nylon membranes, which are widely available. The semi-permeable membranes may be mounted on a more robust solid surface such as glass. The surfaces may optionally be coated with a layer of metal, such as gold, platinum or other transition metal. A particular example of a suitable solid substrate is the commercially available BIACore™ chip (Pharmacia Biosensors).

Preferably, the solid substrate is generally a material having a rigid or semi-rigid surface. In preferred embodiments, at least one surface of the substrate will be substantially flat, although in some embodiments it may be desirable to physically separate synthesis regions for different polymers with, for example, raised regions or etched trenches. It is also preferred that the solid substrate is suitable for the high density application of DNA sequences in discrete areas of typically from 50 to 100 μm, giving a density of 10000 to 40000 cm^{-2}.

The solid substrate is conveniently divided up into sections. This may be achieved by techniques such as photoetching, or by the application of hydrophobic inks, for example teflon-based inks (Cel-line, USA).

Discrete positions, in which each different member of the array is located may have any convenient shape, e.g., circular, rectangular, elliptical, wedge-shaped, etc.

Attachment of the polynucleotide sequences to the substrate may be by covalent or non-covalent means. The plurality of polynucleotide sequences may be attached to the substrate via a layer of molecules to which the sequences bind. For example, the sequences may be labelled with biotin and the substrate coated with avidin and/or streptavidin. A convenient feature of using biotinylated sequences is that the efficiency of coupling to the solid substrate can be

determined easily. Since the library sequences may bind only poorly to some solid substrates, it is often necessary to provide a chemical interface between the solid substrate (such as in the case of glass) and the sequences. Examples of suitable chemical interfaces include hexaethylene glycol. Another example is the use of polylysine coated glass, the polylysine then being chemically modified using standard procedures to introduce an affinity ligand. Other methods for attaching molecules to the surfaces of solid substrate by the use of coupling agents are known in the art, see for example WO98/49557.

The complete DNA array is typically read at the same time by charged coupled device (CCD) camera or confocal imaging system. Alternatively, the DNA array may be placed for detection in a suitable apparatus that can move in an x-y direction, such as a plate reader. In this way, the change in characteristics for each discrete position can be measured automatically by computer controlled movement of the array to place each discrete element in turn in line with the detection means.

The detection means are capable of interrogating each position in the library array optically or electrically. Examples of suitable detection means include CCD cameras or confocal imaging systems.

Nucleotide vectors

Polynucleotides of the invention can be incorporated into a recombinant replicable vector. The vector may be used to replicate the nucleic acid in a compatible host cell. Thus in a further embodiment, the invention provides a method of making polynucleotides of the invention by introducing a polynucleotide of the invention into a replicable vector, introducing the vector into a compatible host cell, and growing the host cell under conditions which bring about replication of the vector. The vector may be recovered from the host cell. Suitable host cells include bacteria such as *E. coli*, yeast, mammalian cell lines and other eukaryotic cell lines, for example insect Sf9 cells.

Preferably, a polynucleotide of the invention in a vector is operably linked to a control sequence that is capable of providing for the expression of the coding sequence by the host cell, i.e. the vector is an expression vector. The term "operably linked" means that the components described are in a relationship permitting them to function in their intended manner. A regulatory sequence "operably linked" to a coding sequence is ligated in such a way that expression of the coding sequence is achieved under condition compatible with the control sequences.

The control sequences may be modified, for example by the addition of further transcriptional regulatory elements to make the level of transcription directed by the control sequences more responsive to transcriptional modulators.

Vectors of the invention may be transformed or transfected into a suitable host cell as described below to provide for expression of a protein of the invention. This process may comprise culturing a host cell transformed with an expression vector as described above under conditions to provide for expression by the vector of a coding sequence encoding the protein, and optionally recovering the expressed protein.

The vectors may be for example, plasmid or virus vectors provided with an origin of replication, optionally a promoter for the expression of the said polynucleotide and optionally a regulator of the promoter. The vectors may contain one or more selectable marker genes, for example an ampicillin resistance gene in the case of a bacterial plasmid or a neomycin resistance gene for a mammalian vector. Vectors may be used, for example, to transfect or transform a host cell.

Control sequences operably linked to sequences encoding the protein of the invention include promoters/enhancers and other expression regulation signals. These control sequences may be selected to be compatible with the host cell for which the expression vector is designed to be used in. The term "promoter" is well-known in the art and encompasses nucleic acid regions ranging in size and complexity from minimal promoters to promoters including upstream elements and enhancers.

The promoter is typically selected from promoters which are functional in mammalian cells, although prokaryotic promoters and promoters functional in other eukaryotic cells may be used. The promoter is typically derived from promoter sequences of viral or eukaryotic genes. For example, it may be a promoter derived from the genome of a cell in which expression is to occur. With respect to eukaryotic promoters, they may be promoters that function in a ubiquitous manner (such as promoters of alpha-actin, beta-actin, tubulin) or, alternatively, a tissue-specific manner (such as promoters of the genes for pyruvate kinase). Tissue-specific promoters specific for particular cells may be used if appropriate. They may also be promoters that respond to specific stimuli, for example promoters that bind steroid hormone receptors. Viral promoters may also be used, for example the Moloney murine leukaemia virus long terminal repeat (MMLV LTR) promoter, the rous sarcoma virus (RSV) LTR promoter or the human cytomegalovirus (CMV) IE promoter.

It may also be advantageous for the promoters to be inducible so that the levels of expression of the heterologous gene can be regulated during the life-time of the cell. Inducible means that the levels of expression obtained using the promoter can be regulated.

In addition, any of these promoters may be modified by the addition of further regulatory sequences, for example enhancer sequences. Chimeric promoters may also be used comprising sequence elements from two or more different promoters described above.

Host cells

Vectors and polynucleotides of the invention may be introduced into host cells for the purpose of replicating the vectors/polynucleotides and/or expressing the proteins of the invention encoded by the polynucleotides of the invention. Although the proteins of the invention may be produced using prokaryotic cells as host cells, it is preferred to use eukaryotic cells, for example yeast, insect or mammalian cells, in particular mammalian cells.

Vectors/polynucleotides of the invention may introduced into suitable host cells using a variety of techniques known in the art, such as transfection, transformation and electroporation. Where vectors/polynucleotides of the invention are to be administered to animals, several techniques are known in the art, for example infection with recombinant viral vectors such as retroviruses, herpes simplex viruses and adenoviruses, direct injection of nucleic acids and biolistic transformation.

Polypeptides

Polypeptides of the invention comprise polypeptides encoded by polynucleotides of the invention. It will be understood that polypeptide sequences of the invention include sequences homologous to the polypeptides encoded by the nucleotide sequences referred to in the Tables, which may be obtained from any source, for example related viral/bacterial proteins, cellular homologues and synthetic peptides, as well as variants or derivatives thereof.

Thus, the present invention encompasses variants, homologues or derivatives of the amino acid sequences of the present invention, as well as variants, homologues or derivatives of the amino acid sequences coded for by the nucleotide sequences of the present invention. In one embodiment, homologues are naturally occurring sequences, such as orthologues, tissue-specific isoforms and allelic variants.

In the context of the present invention, a homologous sequence is taken to include an amino acid sequence which is at least 60, 70, 80 or 90% identical, preferably at least 95 or 98% identical at the amino acid level over at least 20, 40, 60 or 80 amino acids with a sequence encoded by a nucleotide sequence referred to in the Tables. In particular, homology should typically be considered with respect to those regions of the sequence known to be essential for specific biological functions rather than non-essential neighbouring sequences.

Although amino acid homology can also be considered in terms of similarity (i.e. amino acid residues having similar chemical properties/functions), in the context of the present invention it is preferred to express homology in terms of sequence identity.

Homology comparisons can be carried out as described above for nucleotide sequences with the appropriate modifications for amino acid sequences. For example when using the GCG Wisconsin Bestfit package (see below) the default gap penalty for amino acid sequences is -12 for a gap and -4 for each extension.

It should also be noted that where computer algorithms are used to align amino acid sequences, although the final % homology can be measured in terms of identity, the alignment process itself is typically not based on an all-or-nothing pair comparison. Instead, a scaled similarity score matrix is generally used that assigns scores to each pairwise comparison based on chemical similarity or evolutionary distance. An example of such a matrix commonly used is the BLOSUM62 matrix - the default matrix for the BLAST suite of programs. GCG Wisconsin programs generally use either the public default values or a custom symbol comparison table if supplied (see user manual for further details). It is preferred to use the public default values for the GCG package, or in the case of other software, the default matrix, such as BLOSUM62.

The terms "variant" or "derivative" in relation to the amino acid sequences of the present invention includes any substitution of, variation of, modification of, replacement of, deletion of or addition of one (or more) amino acids from or to the sequence providing the resultant amino acid sequence preferably has biological activity, preferably having at least 25 to 50% of the activity as the polypeptides referred to in the sequence listings, more preferably at least substantially the same activity. Particular details of biological activity for each polypeptide are given in the Tables.

Thus, the polypeptides referred to in the Tables and homologues thereof, may be modified for use in the present invention. Typically, modifications are made that maintain the activity of the sequence. Thus, in one embodiment, amino acid substitutions may be made, for example from 1, 2 or 3 to 10, 20 or 30

substitutions provided that the modified sequence retains at least about 25 to 50% of, or substantially the same activity. However, in an alternative embodiment, modifications to the amino acid sequences of a polypeptide of the invention may be made intentionally to reduce the biological activity of the polypeptide. For example truncated polypeptides that remain capable of binding to target molecules but lack functional effector domains may be useful as inhibitors of the biological activity of the full length molecule.

In general, preferably less than 20%, 10% or 5% of the amino acid residues of a variant or derivative are altered as compared with the corresponding region of the polypeptides referred to in the Tables.

Amino acid substitutions may include the use of non-naturally occurring analogues, for example to increase blood plasma half-life of a therapeutically administered polypeptide (see below for further details on the production of peptide derivatives for use in therapy).

Conservative substitutions may be made, for example according to the Table below. Amino acids in the same block in the second column and preferably in the same line in the third column may be substituted for each other:

ALIPHATIC	Non-polar	G A P
		I L V
	Polar - uncharged	C S T M
		N Q
	Polar - charged	D E
		K R
AROMATIC		H F W Y

Polypeptides of the invention also include fragments of the above mentioned full length polypeptides and variants thereof, including fragments of the sequences referred to in the Tables and homologues thereof. Preferred fragments include those which include an epitope. Suitable fragments will be at least about 6 or 8, e.g. at least 10, 12, 15 or 20 amino acids in length. They may also be less than 200, 100 or 50 amino acids in length. Polypeptide fragments may contain one or more (e.g. 2, 3, 5, or 10) substitutions, deletions or insertions, including conserved substitutions. Where substitutions, deletion and/or insertions have been made, for example by means of recombinant technology, preferably less than 20%, 10% or 5% of the amino acid residues depicted in the sequence listings are altered.

Proteins of the invention are typically made by recombinant means, for example as described below. However they may also be made by synthetic means using techniques well known to skilled persons such as solid phase synthesis. Various techniques for chemically synthesising peptides are reviewed by Borgia and Fields, 2000, TibTech 18: 243-251 and described in detail in the references contained therein.

Proteins of the invention may also be produced as fusion proteins, for example to aid in extraction and purification. Examples of fusion protein partners include glutathione-S-transferase (GST), hexahistidine, GAL4 (DNA binding and/or transcriptional activation domains) and β-galactosidase. It may also be convenient to include a proteolytic cleavage site between the fusion protein partner and the protein sequence of interest to allow removal of fusion protein sequences. Preferably the fusion protein will not hinder the function of the protein of interest sequence.

Proteins of the invention may be in a substantially isolated form. It will be understood that the protein may be mixed with carriers or diluents which will not interfere with the intended purpose of the protein and still be regarded as substantially isolated. A protein of the invention may also be in a substantially purified form, in which case it will generally comprise the protein in a preparation in which more than 90%, e.g. 95%, 98% or 99% of the protein in the preparation is a protein of the invention.

Characteristics of ovarian cancer-associated proteins

Ovarian cancer proteins of the present invention may be classified as secreted proteins, transmembrane proteins or intracellular proteins. In one embodiment, the ovarian cancer protein is an intracellular protein. Intracellular proteins may be found in the cytoplasm and/or in the nucleus. Intracellular proteins are involved in all aspects of cellular function and replication (including, e.g., signaling pathways); aberrant expression of such proteins often results in unregulated or disregulated cellular processes *(see,* e.g., Molecular Biology of the Cell (Alberts, ed., 3rd ed., 1994). For example, many intracellular proteins have enzymatic activity such as protein kinase activity, protein phosphatase activity, protease activity, nucleotide cyclase activity, polymerase activity and the like. Intracellular proteins also serve as docking proteins that are involved in organizing complexes of proteins, or targeting proteins to various subcellular localizations, and are involved in maintaining the structural integrity of organelles.

An increasingly appreciated concept in characterising proteins is the presence in the proteins of one or more motifs for which defined functions have

been attributed. In addition to the highly conserved sequences found in the enzymatic domain of proteins, highly conserved sequences have been identified in proteins that are involved in protein-protein interaction. For example, Src-homology-2 (SH2) domains bind tyrosine-phosphorylated targets in a sequence dependent manner. PTB domains, which are distinct from SH2 domains, also bind tyrosine phosphorylated targets. SH3 domains bind to proline-rich targets. In addition, PH domains, tetratricopeptide repeats and WD domains to name only a few, have been shown to mediate protein-protein interactions. Some of these may also be involved in binding to phospholipids or other second messengers. As will be appreciated by one of ordinary skill in the art, these motifs can be identified on the basis of primary sequence; thus, an analysis of the sequence of proteins may provide insight into both the enzymatic potential of the molecule and/or molecules with which the protein may associate. One useful database is Pfam (protein families), which is a large collection of multiple sequence alignments and hidden Markov models covering many common protein domains. Versions are available via the internet from Washington University in St. Louis, the Sanger Center in England, and the Karolinska Institute in Sweden (*see*, e.g., Bateman et al., 2000, Nuc. Acids Res. 28: 263-266; Sonnhammer et al., 1997, Proteins 28: 405-420; Bateman et al., 1999, Nuc. Acids Res. 27:260-262; and Sonnhammer et al., 1998, Nuc. Acids Res. 26: 320-322.

In another embodiment, the ovarian cancer sequences are transmembrane proteins. Transmembrane proteins are molecules that span a phospholipid bilayer of a cell. They may have an intracellular domain, an extracellular domain, or both. The intracellular domains of such proteins may have a number of functions including those already described for intracellular proteins. For example, the intracellular domain may have enzymatic activity and/or may serve as a binding site for additional proteins. Frequently the intracellular domain of transmembrane proteins serves both roles. For example certain receptor tyrosine kinases have both protein kinase activity and SH2 domains. In addition, autophosphorylation of tyrosines on the receptor molecule itself, creates binding sites for additional SH2 domain containing proteins.

Transmembrane proteins may contain from one to many transmembrane domains. For example, receptor tyrosine kinases, certain cytokine receptors, receptor guanylyl cyclases and receptor serine/threonine protein kinases contain a single transmembrane domain. However, various other proteins including channels and adenylyl cyclases contain numerous transmembrane domains. Many important cell surface receptors such as G protein coupled receptors

(GPCRs) are classified as "seven transmembrane domain" proteins, as they contain 7 membrane spanning regions. Characteristics of transmembrane domains include approximately 20 consecutive hydrophobic amino acids that may be followed by charged amino acids. Therefore, upon analysis of the amino acid sequence of a particular protein, the localization and number of transmembrane domains within the protein may be predicted *(see, e.g.* PSORT web site http://psort.nibb.ac.jp/). Important transmembrane protein receptors include, but are not limited to the insulin receptor, insulin-like growth factor receptor, human growth hormone receptor, glucose transporters, transferrin receptor, epidermal growth factor receptor, low density lipoprotein receptor, epidermal growth factor receptor, leptin receptor, interleukin receptors, e.g. IL-1 receptor, IL-2 receptor,

The extracellular domains of transmembrane proteins are diverse; however, conserved motifs are found repeatedly among various extracellular domains. Conserved structure and/or functions have been ascribed to different extracellular motifs. Many extracellular domains are involved in binding to other molecules. In one aspect, extracellular domains are found on receptors. Factors that bind the receptor domain include circulating ligands, which may be peptides, proteins, or small molecules such as adenosine and the like. For example, growth factors such as EGF, FGF and PDGF are circulating growth factors that bind to their cognate receptors to initiate a variety of cellular responses. Other factors include cytokines, mitogenic factors, neurotrophic factors and the like. Extracellular domains also bind to cell-associated molecules. In this respect, they mediate cell-cell interactions., Cell-associated ligands can be tethered to the cell, e.g., via a glycosylphosphatidylinositol (GPI) anchor, or may themselves be transmembrane proteins. Extracellular domains also associate with the extracellular matrix and contribute to the maintenance of the cell structure.

Ovarian cancer proteins that are transmembrane are particularly preferred in the present invention as they are readily accessible targets for immunotherapeutics, as are described herein. In addition, as outlined below, transmembrane proteins can be also useful in imaging modalities. Antibodies may be used to label such readily accessible proteins *in situ.* Alternatively, antibodies can also label intracellular proteins, in which case samples are typically permeablized to provide access to intracellular proteins.

It will also be appreciated by those in the art that a transmembrane protein can be made soluble by removing transmembrane sequences, e.g., through recombinant methods. Furthermore, transmembrane proteins that have been made soluble can be made to be secreted through recombinant means by adding an appropriate signal sequence.

In another embodiment, the ovarian cancer proteins are secreted proteins; the secretion of which can be either constitutive or regulated. These proteins have a signal peptide or signal sequence that targets the molecule to the secretory pathway. Secreted proteins are involved in numerous physiological events; by virtue of their circulating nature, they serve to transmit signals to various other cell types. The secreted protein may function in an autocrine manner (acting on the cell that secreted the factor), a paracrine manner (acting on cells in close proximity to the cell that secreted the factor) or an endocrine manner (acting on cells at a distance). Thus secreted molecules find use in modulating or altering numerous aspects of physiology. Ovarian cancer proteins that are secreted proteins are particularly preferred in the present invention as they serve as good targets for diagnostic markers, e.g., for blood, plasma, serum, or stool tests.

Therapeutic peptides

Polypeptides/peptides of the present invention may be administered therapeutically to patients. It is preferred to use peptides that do not consisting solely of naturally-occurring amino acids but which have been modified, for example to reduce immunogenicity, to increase circulatory half-life in the body of the patient, to enhance bioavailability and/or to enhance efficacy and/or specificity.

A number of approaches have been used to modify peptides for therapeutic application. One approach is to link the peptides or proteins to a variety of polymers, such as polyethylene glycol (PEG) and polypropylene glycol (PPG) – see for example U.S. Patent Nos. 5,091,176, 5,214,131 and US 5,264,209.

Replacement of naturally-occurring amino acids with a variety of uncoded or modified amino acids such as D-amino acids and N-methyl amino acids may also be used to modify peptides

Another approach is to use bifunctional crosslinkers, such as N-succinimidyl 3-(2 pyridyldithio) propionate, succinimidyl 6-[3-(2 pyridyldithio) propionamido] hexanoate, and sulfosuccinimidyl 6-[3-(2 pyridyldithio) propionamido]hexanoate (see US Patent 5,580,853).

It may be desirable to use derivatives of the peptides of the invention which are conformationally constrained. Conformational constraint refers to the stability and preferred conformation of the three-dimensional shape assumed by a peptide. Conformational constraints include local constraints, involving restricting the conformational mobility of a single residue in a peptide; regional constraints,

involving restricting the conformational mobility of a group of residues, which residues may form some secondary structural unit; and global constraints, involving the entire peptide structure.

The active conformation of the peptide may be stabilized by a covalent modification, such as cyclization or by incorporation of gamma-lactam or other types of bridges. For example, side chains can be cyclized to the backbone so as create a L-gamma-lactam moiety on each side of the interaction site. See, generally, Hruby et al., "Applications of Synthetic Peptides," in Synthetic Peptides: A User's Guide: 259-345 (W. H. Freeman & Co. 1992). Cyclization also can be achieved, for example, by formation of cystine bridges, coupling of amino and carboxy terminal groups of respective terminal amino acids, or coupling of the amino group of a Lys residue or a related homolog with a carboxy group of Asp, Glu or a related homolog. Coupling of the .alpha-amino group of a polypeptide with the epsilon-amino group of a lysine residue, using iodoacetic anhydride, can be also undertaken. See Wood and Wetzel, 1992, Int'l J. Peptide Protein Res. 39: 533-39.

Another approach described in US 5,891,418 is to include a metal-ion complexing backbone in the peptide structure. Typically, the preferred metal-peptide backbone is based on the requisite number of particular coordinating groups required by the coordination sphere of a given complexing metal ion. In general, most of the metal ions that may prove useful have a coordination number of four to six. The nature of the coordinating groups in the peptide chain includes nitrogen atoms with amine, amide, imidazole, or guanidino functionalities; sulfur atoms of thiols or disulfides; and oxygen atoms of hydroxy, phenolic, carbonyl, or carboxyl functionalities. In addition, the peptide chain or individual amino acids can be chemically altered to include a coordinating group, such as for example oxime, hydrazino, sulfhydryl, phosphate, cyano, pyridino, piperidino, or morpholino. The peptide construct can be either linear or cyclic, however a linear construct is typically preferred. One example of a small linear peptide is Gly-Gly-Gly-Gly which has four nitrogens (an N_4 complexation system) in the back bone that can complex to a metal ion with a coordination number of four.

A further technique for improving the properties of therapeutic peptides is to use non-peptide peptidomimetics. A wide variety of useful techniques may be used to elucidating the precise structure of a peptide. These techniques include amino acid sequencing, x-ray crystallography, mass spectroscopy, nuclear magnetic resonance spectroscopy, computer-assisted molecular modeling, peptide mapping, and combinations thereof. Structural analysis of a peptide generally provides a large body of data which comprise the amino acid sequence

of the peptide as well as the three-dimensional positioning of its atomic components. From this information, non-peptide peptidomimetics may be designed that have the required chemical functionalities for therapeutic activity but are more stable, for example less susceptible to biological degradation. An example of this approach is provided in US 5,811,512.

Techniques for chemically synthesising therapeutic peptides of the invention are described in the above references and also reviewed by Borgia and Fields, 2000, TibTech 18: 243-251 and described in detail in the references contained therein.

Protein Expression and Purification

Host cells comprising polynucleotides of the invention may be used to express proteins of the invention. Host cells may be cultured under suitable conditions which allow expression of the proteins of the invention. Expression of the proteins of the invention may be constitutive such that they are continually produced, or inducible, requiring a stimulus to initiate expression. In the case of inducible expression, protein production can be initiated when required by, for example, addition of an inducer substance to the culture medium, for example dexamethasone or IPTG.

Proteins of the invention can be extracted from host cells by a variety of techniques known in the art, including enzymatic, chemical and/or osmotic lysis and physical disruption.

Polypeptides of the invention may also be produced recombinantly in an *in vitro* cell-free system, such as the TnT™ (Promega) rabbit reticulocyte system or a wheat germ lysate system.

Antibodies

The invention also provides monoclonal or polyclonal antibodies that bind specifically to polypeptides of the invention or fragments thereof. Thus, the present invention further provides a process for the production of monoclonal or polyclonal antibodies to polypeptides of the invention.

The phrase "binds specifically" to a polypeptide means that the binding of the antibody to the protein or peptide is determinative of the presence of the protein, in a heterogeneous population of proteins and other biologics. Thus, under designated immunoassay conditions, the specified antibodies bind to a particular protein at least two times the background and more typically more than 10 to 100 times background. Typically, antibodies of the invention bind to a protein

of interest with a Kd of at least about 0.1 mM, more usually at least about 1 μM, preferably at least about 0.1 μM, and most preferably at least, 0.01 μM.

If polyclonal antibodies are desired, a selected mammal (e.g., mouse, rabbit, goat, horse, etc.) is immunised with an immunogenic polypeptide bearing a epitope(s). Serum from the immunised animal is collected and treated according to known procedures. If serum containing polyclonal antibodies to a particular polypeptide epitope of interest contains antibodies to other antigens, the polyclonal antibodies can be purified by immunoaffinity chromatography. Techniques for producing and processing polyclonal antisera are known in the art. In order that such antibodies may be made, the invention also provides polypeptides of the invention or fragments thereof haptenised to another polypeptide for use as immunogens in animals or humans.

Monoclonal antibodies directed against particular epitopes in the polypeptides of the invention can also be readily produced by one skilled in the art. The general methodology for making monoclonal antibodies by hybridomas is well known. Immortal antibody-producing cell lines can be created by cell fusion, and also by other techniques such as direct transformation of B lymphocytes with oncogenic DNA, or transfection with Epstein-Barr virus. Panels of monoclonal antibodies produced against particular epitopes can be screened for various properties; i.e., for isotype and epitope affinity.

An alternative technique involves screening phage display libraries where, for example the phage express scFv fragments on the surface of their coat with a large variety of complementarity determining regions (CDRs). This technique is well known in the art.

Antibodies, both monoclonal and polyclonal, which are directed against polypeptides of the invention or fragments thereof are particularly useful in diagnosis, and those which are neutralising are useful in passive immunotherapy. Monoclonal antibodies, in particular, may be used to raise anti-idiotype antibodies. Anti-idiotype antibodies are immunoglobulins which carry an "internal image" of the antigen of the agent against which protection is desired.

Techniques for raising anti-idiotype antibodies are known in the art. These anti-idiotype antibodies may also be useful in therapy.

For the purposes of this invention, the term "antibody", unless specified to the contrary, includes fragments of whole antibodies which retain their binding activity for a target antigen. Such fragments include Fv, F(ab') and $F(ab')_2$ fragments, as well as single chain antibodies (scFv). Furthermore, the antibodies and fragments thereof may be humanised antibodies, for example as described in EP-A-239400.

Antibodies may be used in method of detecting polypeptides of the invention present in biological samples by a method which comprises:

(a) providing an antibody of the invention;

(b) incubating a biological sample with said antibody under conditions which allow for the formation of an antibody-antigen complex; and

(c) determining whether an antibody-antigen complex comprising said antibody is formed.

Such tests for detecting polypeptides of the invention in a biological sample may be used as part of the methods of the invention for detecting ovarian cancer-associated polypeptides and monitoring the efficacy of treatment of patients suffering from ovarian cancer as will be described in more detail below.

Antibodies of the invention may be bound to a solid support and/or packaged into kits in a suitable container along with suitable reagents, controls, instructions and the like.

Construction of Transgenic Animals Expressing Polypeptides of the Invention and "Knock-Out" Animals

The present invention also provides transgenic animals which are transgenic by virtue of comprising a polynucleotide of the invention, i.e. animals transformed with a polynucleotide of the invention. Suitable animals are generally from the phylum chordata. Chordates includes vertebrate groups such as mammals, birds, reptiles and amphibians. Particular examples of mammals include non-human primates, cats, dogs, ungulates such as cows, goats, pigs, sheep and horses and rodents such as mice, rats, gerbils and hamsters. Transgenic animals within the meaning of the present invention are non-human animals and the production of transgenic humans is specifically excluded.

Techniques for producing transgenic animals are well known in the art. A useful general textbook on this subject is Houdebine, Transgenic animals – Generation and Use (Harwood Academic, 1997) – an extensive review of the techniques used to generate transgenic animals from fish to mice and cows.

Advances in technologies for embryo micromanipulation now permit introduction of heterologous DNA into, for example, fertilized mammalian ova. For instance, totipotent or pluripotent stem cells can be transformed by microinjection, calcium phosphate mediated precipitation, liposome fusion, retroviral infection or other means, the transformed cells are then introduced into the embryo, and the embryo then develops into a transgenic animal. In a highly preferred method, developing embryos are infected with a retrovirus containing the desired DNA,

and transgenic animals produced from the infected embryo. In a most preferred method, however, the appropriate DNAs are coinjected into the pronucleus or cytoplasm of embryos, preferably at the single cell stage, and the embryos allowed to develop into mature transgenic animals. Those techniques as well known. See reviews of standard laboratory procedures for microinjection of heterologous DNAs into mammalian fertilized ova, including Hogan *et al.*, Manipulating the Mouse Embryo, (Cold Spring Harbor Press 1986); Krimpenfort *et al.*, Bio/Technology 9:844 (1991); Palmiter *et al.*, Cell, 41: 343 (1985); Kraemer *et al.*, Genetic manipulation of the Mammalian Embryo, (Cold Spring Harbor Laboratory Press 1985); Hammer *et al.*, Nature, 315: 680 (1985); Wagner *et al.*, U.S. Pat. No. 5,175,385; Krimpenfort *et al.*, U.S. Pat. No. 5,175,384, the respective contents of which are incorporated herein by reference

Another method used to produce a transgenic animal involves microinjecting a nucleic acid into pro-nuclear stage eggs by standard methods. Injected eggs are then cultured before transfer into the oviducts of pseudopregnant recipients.

Transgenic animals may also be produced by nuclear transfer technology as described in Schnieke, A.E. *et al.*, 1997, Science, 278: 2130 and Cibelli, J.B. *et al.*, 1998, Science, 280: 1256. Using this method, fibroblasts from donor animals are stably transfected with a plasmid incorporating the coding sequences for a binding domain or binding partner of interest under the control of regulatory. Stable transfectants are then fused to enucleated oocytes, cultured and transferred into female recipients.

Analysis of animals which may contain transgenic sequences would typically be performed by either PCR or Southern blot analysis following standard methods.

By way of a specific example for the construction of transgenic mammals, such as cows, nucleotide constructs comprising a sequence encoding a binding domain fused to GFP are microinjected using, for example, the technique described in U.S. Pat. No. 4,873,191, into oocytes which are obtained from ovaries freshly removed from the mammal. The oocytes are aspirated from the follicles and allowed to settle before fertilization with thawed frozen sperm capacitated with heparin and prefractionated by Percoll gradient to isolate the motile fraction.

The fertilized oocytes are centrifuged, for example, for eight minutes at 15,000 g to visualize the pronuclei for injection and then cultured from the zygote to morula or blastocyst stage in oviduct tissue-conditioned medium. This medium is prepared by using luminal tissues scraped from oviducts and diluted in culture

medium. The zygotes must be placed in the culture medium within two hours following microinjection.

Oestrous is then synchronized in the intended recipient mammals, such as cattle, by administering coprostanol. Oestrous is produced within two days and the embryos are transferred to the recipients 5-7 days after estrous. Successful transfer can be evaluated in the offspring by Southern blot.

Alternatively, the desired constructs can be introduced into embryonic stem cells (ES cells) and the cells cultured to ensure modification by the transgene. The modified cells are then injected into the blastula embryonic stage and the blastulas replaced into pseudopregnant hosts. The resulting offspring are chimeric with respect to the ES and host cells, and nonchimeric strains which exclusively comprise the ES progeny can be obtained using conventional cross-breeding. This technique is described, for example, in WO91/10741.

In another embodiment, transgenic animals of the present invention are transgenic "knock-out" animals where a specific gene corresponding to a polynucleotide referred to in the Tables has been rendered non-functional by homologous recombination. The generation of "knock-out" animals is similar to the production of other transgenic animals except that the polynucleotide constructs are designed to integrate into the endogenous genes and disrupt the function of the endogenous sequences. The generation of "knock-out" animals is known in the art, including the design of suitable constructs that will recombine at the appropriate site in the genome.

In one embodiment, the heterologous sequence which it is desired to recombine into the genome of a target animal comprises a functional sequence but under the control of an inducible promoter so that expression of the gene can be regulated by administration of an endogenous molecule. This may be advantageous where disruption of the gene is embryonic-lethal.

"Knock-out" animals can be used as animal models for the study of gene function.

Informatics

The ability to identify genes that are over or under expressed in ovarian cancer can additionally provide high-resolution, high-sensitivity datasets which can be used in the areas of diagnostics, therapeutics, drug development, pharmacogenetics, protein structure, biosensor development, and other related areas. For example, the expression profiles can be used in diagnostic or prognostic evaluation of patients with ovarian cancer. Or as another example, subcellular toxicological information can be generated to better direct drug

structure and activity correlation *(see* Anderson, *Pharmaceutical Proteomics: Targets, Mechanism, and Function,* paper presented at the IBC Proteomics conference, Coronado, CA (June 11-12, 1998)). Subcellular toxicological information can also be utilized in a biological sensor device to predict the likely toxicological effect of chemical exposures and likely tolerable exposure thresholds *(see* U.S. Patent No. 5,811,231). Similar advantages accrue from datasets relevant to other biomolecules and bioactive agents (e.g., nucleic acids, saccharides, lipids, drugs, and the like).

Thus, in another embodiment, the present invention provides a database that includes at least one set of assay data. The data contained in the database is acquired, e.g., using array analysis either singly or in a library format. The database can be in substantially any form in which data can be maintained and transmitted, but is preferably an electronic database. The electronic database of the invention can be maintained on any electronic device allowing for the storage of and access to the database, such as a personal computer, but is preferably distributed on a wide area network, such as the World Wide Web.

The focus of the present section on databases that include peptide sequence data is for clarity of illustration only. It will be apparent to those of skill in the art that similar databases can be assembled for any assay data acquired using an assay of the invention.

The compositions and methods for identifying and/or quantitating the relative and/or absolute abundance of a variety of molecular and macromolecular species from a biological sample undergoing ovarian cancer, i.e., the identification of ovarian cancer-associated sequences described herein, provide an abundance of information, which can be correlated with pathological conditions, predisposition to disease, drug testing, therapeutic monitoring, gene-disease causal linkages, identification of correlates of immunity and physiological status, among others. Although the data generated from the assays of the invention is suited for manual review and analysis, in a preferred embodiment, prior data processing using high-speed computers is utilized.

An array of methods for indexing and retrieving biomolecular information is known in the art. For example, U.S. Patents 6,023,659 and 5,966,712 disclose a relational database system for storing biomolecular sequence information in a manner that allows sequences to be catalogued and searched according to one or more protein function hierarchies. U.S. Patent 5,953,727 discloses a relational database having sequence records containing information in a format that allows a collection of partial-length DNA sequences to be catalogued and searched according to association with one or more sequencing projects for obtaining

full-length sequences from the collection of partial length sequences. U.S. Patent 5,706,498 discloses a gene database retrieval system for making a retrieval of a gene sequence similar to a sequence data item in a gene database based on the degree of similarity between a key sequence and a target sequence. U.S. Patent 5,538,897 discloses a method using mass spectroscopy fragmentation patterns of peptides to identify amino acid sequences in computer databases by comparison of predicted mass spectra with experimentally-derived mass spectra using a closeness-of-fit measure. U.S. Patent 5,926,818 discloses a multi-dimensional database comprising a functionality for multi-dimensional data analysis described as on-line analytical processing (OLAP), which entails the consolidation of projected and actual data according to more than one consolidation path or dimension. U.S. Patent 5,295,261 reports a hybrid database structure in which the fields of each database record are divided into two classes, navigational and informational data, with navigational fields stored in a hierarchical topological map which can be viewed as a tree structure or as the merger of two or more such tree structures.

See also Mount *et al., Bioinformatics (2001); Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids* (Durbin *et al.*, eds., 1999); *Bioiraformatics: A Practical Guide to the Analysis of Genes and Proteins* (Baxevanis & Oeullette eds., 1998)); Rashidi & Buehler, *Bioinformatics: Basic Applications in Biological Science and Medicine* (1999); *Introduction to Computational Molecular Biology* (Setubal *et al.*, eds 1997); *Bioinformatics: Methods and Protocols* (Misener & Krawetz, eds, *2000); Bioinformatics: Sequence, Structure, and Databanks: A Practical Approach* (Higgins & Taylor, eds., 2000); Brown, *Bioinfor7natics: A Biologist's Guide to Biocomputing and the Internet (2001);* Han & Kamber, *Data Mining: Concepts and Techniques (2000);* and Waterman, *Introduction to Computational Biology: Maps, Sequences, and Genomes* (1995).

The present invention provides a computer database comprising a computer and software for storing in computer-retrievable form assay data records cross-tabulated, e.g., with data specifying the source of the target-containing sample from which each sequence specificity record was obtained.

In an exemplary embodiment, at least one of the sources of target-containing sample is from a control tissue sample known to be free of pathological disorders. In a variation, at least one of the sources is a known pathological tissue specimen, e.g., a neoplastic lesion or another tissue specimen to be analyzed for prostate cancer. In another variation, the assay records

cross-tabulate one or more of the following parameters for each target species in a sample: (1) a unique identification code, which can include, e.g., a target molecular structure and/or characteristic separation coordinate (e.g., electrophoretic coordinates); (2) sample source; and (3) absolute and/or relative quantity of the target species present in the sample.

The invention also provides for the storage and retrieval of a collection of target data in a computer data storage apparatus, which can include magnetic disks, optical disks, magneto-optical disks, DRAM, SRAM, SGRAM, SDRAM, RDRAM, DDR RAM, magnetic bubble memory devices, and other data storage devices, including CPU registers and on-CPU data storage arrays. Typically, the target data records are stored as a bit pattern in an array of magnetic domains on a magnetizable medium or as an array of charge states or transistor gate states, such as an array of cells in a DRAM device (e.g., each cell comprised of a transistor and a charge storage area, which may be on the transistor). In one embodiment, the invention provides such storage devices, and computer systems built therewith, comprising a bit pattern encoding a protein expression fingerprint record comprising unique identifiers for at least 10 target data records cross-tabulated with target source.

When the target is a peptide or nucleic acid, the invention preferably provides a method for identifying related peptide or nucleic acid sequences, comprising performing a computerised comparison between a peptide or nucleic acid sequence assay record stored in or retrieved from a computer storage device or database and at least one other sequence. The comparison can include a sequence analysis or comparison algorithm or computer program embodiment thereof (e.g., BLAST, FASTA, TFASTA, GAP, BESTFIT – see above) and/or the comparison may be of the relative amount of a peptide or nucleic acid sequence in a pool of sequences determined from a polypeptide or nucleic acid sample of a specimen.

The invention also preferably provides a magnetic disk, such as an IBM-compatible (DOS, Windows, Windows95/.98/2000, Windows NT, OS/2) or other format (e.g., Linux, SunOS, Solaris, AIX, SCO Unix, VMS, MV, Macintosh, etc.) floppy diskette or hard (fixed, Winchester) disk drive, comprising a bit pattern encoding data from an assay of the invention in a file format suitable for retrieval and processing in a computerized sequence analysis, comparison, or relative quantitation method.

The invention also provides a network, comprising a plurality of computing devices linked via a data link, such as an Ethernet cable (coax or IOBaseT), telephone line, ISDN line, wireless network, optical fiber, or other suitable signal

transmission medium, whereby at least one network device (e.g., computer, disk array, etc.) comprises a pattern of magnetic domains (e.g., magnetic disk) and/or charge domains (e.g., an array of DRAM cells) composing a bit pattern encoding data acquired from an assay of the invention.

The invention also provides a method for transmitting assay data that includes generating an electronic signal on an electronic communications device, such as a modem, ISDN terminal adapter, DSL, cable modem, ATM switch, or the like, wherein the signal includes (in native or encrypted format) a bit pattern encoding data from an assay or a database comprising a plurality of assay results obtained by the method of the invention.

In a preferred embodiment, the invention provides a computer system for comparing a query target to a database containing an array of data structures, such as an assay result obtained by the method of the invention, and ranking database targets based on the degree of identity and gap weight to the target data. A central processor is preferably initialized to load and execute the computer program for alignment and/or comparison of the assay results. Data for a query target is entered into the central processor via an I/O device. Execution of the computer program results in the central processor retrieving the assay data from the data file, which comprises a binary description of an assay result.

The target data or record and the computer program can be transferred to secondary memory, which is typically random access memory (e.g., DRAM, SRAM, SGRAM, or SDRAM). Targets are ranked according to the degree of correspondence between a selected assay characteristic (e.g., binding to a selected affinity moiety) and the same characteristic of the query target and results are output via an I/O device. For example, a central processor can be a conventional computer (e.g., Intel Pentium, PowerPC, Alpha, PA-8000, SPARC, MIPS 4400, MIPS 10000, VAX, etc.); a program can be a commercial or public domain molecular biology software package (e.g., UWGCG Sequence Analysis Software, Darwin); a data file can be an optical or magnetic disk, a data server, a memory device (e.g., DRAM, SRAM, SGRAM, SDRAM, EPROM, bubble memory, flash memory, etc.); an I/O device can be a terminal comprising a video display and a keyboard, a modem, an ISDN terminal adapter, an Ethernet port, a punched card reader, a magnetic strip reader, or other suitable I/O device.

The invention also preferably provides the use of a computer system, such as that described above, which comprises: (1) a computer; (2) a stored bit pattern encoding a collection of peptide sequence specificity records obtained by the methods of the invention, which may be stored in the computer; (3) a comparison target, such as a query target; and (4) a program for alignment and comparison,

typically with rank-ordering of comparison results on the basis of computed similarity values.

Detection of ovarian cancer sequence for diagnostic and therapeutic applications

In one aspect, the RNA expression levels of genes are determined for different cellular states in the ovarian cancer phenotype. Expression levels of genes in normal tissue (i.e., not undergoing ovarian cancer) and in ovarian cancer tissue (and in some cases, for varying severities of ovarian cancer that relate to prognosis, as outlined below) are evaluated to provide expression profiles. An expression profile of a particular cell state or point of development is essentially a "fingerprint" of the state. While two states may have any particular gene similarly expressed, the evaluation of a number of genes simultaneously allows the generation of a gene expression profile that is reflective of the state of the cell. By comparing expression profiles of cells in different states, information regarding which genes are important (including both up- and down-regulation of genes) in each of these states is obtained. Then, diagnosis may be performed or confirmed to determine whether a tissue sample has the gene expression profile of normal or cancerous tissue. This will provide for molecular diagnosis of related conditions.

"Differential expression," or grammatical equivalents as used herein, refers to qualitative or quantitative differences in the temporal and/or cellular gene expression patterns within and among cells and tissue. Thus, a differentially expressed gene can qualitatively have its expression altered, including an activation or inactivation, in, e.g., normal versus ovarian cancer tissue. Genes may be turned on or turned off in a particular state, relative to another state thus permitting comparison of two or more states. A qualitatively regulated gene will exhibit an expression pattern within a state or cell type which is detectable by standard techniques. Some genes will be expressed in one state or cell type, but not in both. Alternatively, the difference in expression may be quantitative, e.g., in that expression is increased or decreased; i.e., gene expression is either upregulated, resulting in an increased amount of transcript, or downregulated, resulting in a decreased amount of transcript. The degree to which expression differs need only be large enough to quantify via standard characterization techniques as outlined below, such as by use of Affymetrix GeneChipTM expression arrays, Lockhart, *Nature Biotechnology 14:1675-1680 (1996),* hereby expressly incorporated by reference. Other techniques include, but are not limited to, quantitative reverse transcriptase PCR, northern analysis and RNase

protection. As outlined above, preferably the change in expression (i.e., upregulation or downregulation) is at least about 50%, more preferably at least about 100%, more preferably at least about 150%, more preferably at least about 200%, with from 300 to at least 1000% being especially preferred.

Evaluation may be at the gene transcript, or the protein level. The amount of gene expression may be monitored using nucleic acid probes to the DNA or RNA equivalent of the gene transcript, and the quantification of gene expression levels, or, alternatively, the final gene product itself (protein) can be monitored, e.g., with antibodies to the ovarian cancer protein and standard immunoassays (ELISAs, etc.) or other techniques, including mass spectroscopy assays, 2D gel electrophoresis assays, etc. Proteins corresponding to ovarian cancer genes, i.e., those identified as being important in a ovarian cancer phenotype, can be evaluated in a ovarian cancer diagnostic test.

In a preferred embodiment, gene expression monitoring is performed on a plurality of genes. Multiple protein expression monitoring can be performed as well. Similarly, these assays may be performed on an individual basis as well.

In this embodiment, the ovarian cancer nucleic acid probes are attached to biochips as outlined herein for the detection and quantification of ovarian cancer sequences in a particular cell. The assays are further described below in the example. PCR techniques can be used to provide greater sensitivity.

In a preferred embodiment nucleic acids encoding the ovarian cancer protein are detected. Although DNA or RNA encoding the ovarian cancer protein may be detected, of particular interest are methods wherein an mRNA encoding a ovarian cancer protein is detected. Probes to detect mRNA can be a nucleotide/deoxynucleotide probe that is complementary to and hybridizes with the mRNA and includes, but is not limited to, oligonucleotides, cDNA or RNA. Probes also should contain a detectable label, as defined herein. In one method the mRNA is detected after immobilizing the nucleic acid to be examined on a solid support such as nylon membranes and hybridizing the probe with the sample. Following washing to remove the non-specifically bound probe, the label is detected. In another method detection of the mRNA is performed in situ. In this method permeabilized cells or tissue samples are contacted with a detectably labeled nucleic acid probe for sufficient time to allow the probe to hybridize with the target mRNA. Following washing to remove the non-specifically bound probe, the label is detected. For example a digoxygenin labeled riboprobe (RNA probe) that is complementary to the mRNA encoding a ovarian cancer protein is detected by binding the digoxygenin with an anti-digoxygenin secondary antibody

and developed with nitro blue tetrazolium and 5-bromo-4-chloro-3indoyl phosphate.

In a preferred embodiment, various proteins from the three classes of proteins as described herein (secreted, transmembrane or intracellular proteins) are used in diagnostic assays. The ovarian cancer proteins, antibodies, nucleic acids, modified proteins and cells containing ovarian cancer sequences are used in diagnostic assays. This can be performed on an individual gene or corresponding polypeptide level. In a preferred embodiment, the expression profiles are used, preferably in conjunction with high throughput screening techniques to allow monitoring for expression profile genes and/or corresponding polypeptides.

As described and defined herein, ovarian cancer proteins, including intracellular, transmembrane or secreted proteins, find use as markers of ovarian cancer. Detection of these proteins in putative ovarian cancer tissue allows for detection or diagnosis of ovarian cancer. In one embodiment, antibodies are used to detect ovarian cancer proteins. A preferred method separates proteins from a sample by electrophoresis on a gel (typically a denaturing and reducing protein gel, but may be another type of gel, including isoelectric focusing gels and the like). Following separation of proteins, the ovarian cancer protein is detected, e.g., by immunoblotting with antibodies raised against the ovarian cancer protein. Methods of immunoblotting are well known to those of ordinary skill in the art.

In another preferred method, antibodies to the ovarian cancer protein find use in *in situ* imaging techniques, e.g., in histology (e.g., *Methods in Cell Biology: Antibodies in Cell* Biology, volume 37 (Asai, ed. 1993)). In this method cells are contacted with from one to many antibodies to the ovarian cancer protein(s). Following washing to remove non-specific antibody binding, the presence of the antibody or antibodies is detected. In one embodiment the antibody is detected by incubating with a secondary antibody that contains a detectable label. In another method the primary antibody to the ovarian cancer proteins) contains a detectable label, e.g. an enzyme marker that can act on a substrate. In another preferred embodiment each one of multiple primary antibodies contains a distinct and detectable label. This method finds particular use in simultaneous screening for a plurality of ovarian cancer proteins. As will be appreciated by one of ordinary skill in the art, many other histological imaging techniques are also provided by the invention.

In a preferred embodiment the label is detected in a fluorometer which has the ability to detect and distinguish emissions of different wavelengths. In addition, a fluorescence activated cell sorter (FACS) can be used in the method.

In another preferred embodiment, antibodies find use in diagnosing ovarian cancer from blood, serum, plasma, stool, and other samples. Such samples, therefore, are useful as samples to be probed or tested for the presence of ovarian cancer proteins. Antibodies can be used to detect a ovarian cancer protein by previously described immunoassay techniques including ELISA, immunoblotting (western blotting), immunoprecipitation, BIACORE technology and the like. Conversely, the presence of antibodies may indicate an immune response against an endogenous ovarian cancer protein.

In a preferred embodiment, in situ hybridization of labeled ovarian cancer nucleic acid probes to tissue arrays is done. For example, arrays of tissue samples, including ovarian cancer tissue and/or normal tissue, are made. In situ hybridization *(see, e.g.,* Ausubel, supra) is then performed. When comparing the fingerprints between an individual and a standard, the skilled artisan can make a diagnosis, a prognosis, or a prediction based on the findings. It is further understood that the genes which indicate the diagnosis may differ from those which indicate the prognosis and molecular profiling of the condition of the cells may lead to distinctions between responsive or refractory conditions or may be predictive of outcomes.

In a preferred embodiment, the ovarian cancer proteins, antibodies, nucleic acids, modified proteins and cells containing ovarian cancer sequences are used in prognosis assays. As above, gene expression profiles can be generated that correlate to ovarian cancer, in terms of long term prognosis. Again, this may be done on either a protein or gene level, with the use of genes being preferred. As above, ovarian cancer probes may be attached to biochips for the detection and quantification of ovarian cancer sequences in a tissue or patient. The assays proceed as outlined above for diagnosis. PCR method may provide more sensitive and accurate quantification.

Preferred gene sequences, whose expression levels are to be determined as a means of diagnosing ovarian cancer, or a predisposition to ovarian cancer and/or providing long term prognosis are those referred to in the Tables that are listed in one or more of the Tables as having a P value of less than 0.05, more preferably a P value of less than 0.01. Similarly, preferred sequences are those referred to in Table 3B as having a ratio of expression between ovarian patients and normal patients of at least 5, more preferably at least 6, 7, 8, 9 or 10, and those sequences referred to in Table 3B as having a ratio of expression between ovarian patients and normal patients of less than –5 or –10.

The P values and ratios provided in the Tables for the various ovarian cancer-related sequences, including the survival, clinical recurrence and

biochemical recurrence P values in Tables 6, 7 and 8 provide detailed guidance to a person skilled in the art seeking to obtain a diagnosis/prognosis.

Assays for therapeutic compounds

The proteins, nucleic acids, and antibodies as described herein can be used in drug screening assays to identify candidate compounds for use in treating ovarian cancer. The ovarian cancer proteins, antibodies, nucleic acids, modified proteins and cells containing ovarian cancer sequences are used in drug screening assays or by evaluating the effect of drug candidates on a "gene expression profile" or expression profile of polypeptides. In a preferred embodiment, the expression profiles are used, preferably in conjunction with high throughput screening techniques to allow monitoring for expression profile genes after treatment with a candidate agent (e.g., Zlokarnik, *et al.*, 1998, *Science* 279: 84-88); Heid, 1996, *Genome Res* 6: 986-94).

In a preferred embodiment, the ovarian cancer proteins, antibodies, nucleic acids, modified proteins and cells containing the native or modified ovarian cancer proteins are used in screening assays. That is, the present invention provides methods for screening for compounds/agents which modulate the ovarian cancer phenotype or an identified physiological function of a ovarian cancer protein. As above, this can be done on an individual gene level or by evaluating the effect of drug candidates on a "gene expression profile". In a preferred embodiment, the expression profiles are used, preferably in conjunction with high throughput screening techniques to allow monitoring for expression profile genes after treatment with a candidate agent, see Zlokarnik, supra.

Having identified the differentially expressed genes herein, a variety of assays may be executed. In a preferred embodiment, assays may be run on an individual gene or protein level. That is, having identified a particular gene as up regulated in ovarian cancer, test compounds can be screened for the ability to modulate gene expression or for binding to the ovarian cancer protein. "Modulation" thus includes both an increase and a decrease in gene expression. The preferred amount of modulation will depend on the original change of the gene expression in normal versus tissue undergoing ovarian cancer, with changes of at least 10%, preferably 50%, more preferably 100-300%, and in some embodiments 300-1000% or greater. Thus, if a gene exhibits a 4-fold increase in ovarian cancer tissue compared to normal tissue, a decrease of about four-fold is often desired; similarly, a 10-fold decrease in ovarian cancer tissue compared to normal tissue often provides a target value of a 10-fold increase in expression to be induced by the test compound.

The amount of gene expression may be monitored using nucleic acid probes and the quantification of gene expression levels, or, alternatively, the gene product itself can be monitored, e.g., through the use of antibodies to the ovarian cancer protein and standard immunoassays. Proteomics and separation techniques may also allow quantification of expression. .

In a preferred embodiment, gene expression or protein monitoring of a number of entities, i.e., an expression profile, is monitored simultaneously. Such profiles will typically involve a plurality of those entities described herein.

In this embodiment, the ovarian cancer nucleic acid probes are attached to biochips as outlined herein for the detection and quantification of ovarian cancer sequences in a particular cell. Alternatively, PCR may be used. Thus, a series may be used with dispensed primers in desired wells. A PCR reaction can then be performed and analyzed for each well.

Expression monitoring can be performed to identify compounds that modify the expression of one or more ovarian cancer-associated sequences, e.g., a polynucleotide sequence set out in the Tables. In a preferred embodiment, a test modulator is added to the cells prior to analysis. Moreover, screens are also provided to identify agents that modulate ovarian cancer, modulate ovarian cancer proteins, bind to a ovarian cancer protein, or interfere with the binding of a ovarian cancer protein and an antibody or other binding partner.

The term "test compound" or "drug candidate" or "modulator" or grammatical equivalents as used herein describes any molecule, e.g., protein, oligopeptide, small organic molecule, polysaccharide, polynucleotide, etc., to be tested for the capacity to directly or indirectly alter the ovarian cancer phenotype or the expression of a ovarian cancer sequence, e.g., a nucleic acid or protein sequence. In preferred embodiments, modulators alter expression profiles, or expression profile nucleic acids or proteins provided herein. In one embodiment, the modulator suppresses a ovarian cancer phenotype, e.g. to a normal tissue fingerprint. In another embodiment, a modulator induced a ovarian cancer phenotype. Generally, a plurality of assay mixtures are run in parallel with different agent concentrations to obtain a differential response to the various concentrations. Typically, one of these concentrations serves as a negative control, i.e., at zero concentration or below the level of detection.

Drug candidates encompass numerous chemical classes, though typically they are organic molecules, preferably small organic compounds having a molecular weight of more than 100 and less than about 2,500 daltons. Preferred small molecules are less than 2000, or less than 1500 or less than 1000 or less than 500 Daltons. Candidate agents comprise functional groups necessary for

structural interaction with proteins, particularly hydrogen bonding, and typically include at least an amine, carbonyl, hydroxyl or carboxyl group, preferably at least two of the functional chemical groups. The candidate agents often comprise cyclical carbon or heterocyclic structures and/or aromatic or polyaromatic structures substituted with one or more of the above functional groups. Candidate agents are also found among biomolecules including peptides, saccharides, fatty acids, steroids, purines, pyrimidines, derivatives, structural analogs or combinations thereof. Particularly preferred are peptides.

In one aspect, a modulator will neutralize the effect of a ovarian cancer protein. By "neutralize" is meant that activity of a protein is inhibited or blocked and the consequent effect on the cell.

In certain embodiments, combinatorial libraries of potential modulators will be screened for an ability to bind to a ovarian cancer polypeptide or to modulate activity. Conventionally, new chemical entities with useful properties are generated by identifying a chemical compound (called a "lead compound") with some desirable property or activity, e.g., inhibiting activity, creating variants of the lead compound, and evaluating the property and activity of those variant compounds. Often, high throughput screening (HTS) methods are employed for such an analysis.

In one preferred embodiment, high throughput screening methods involve providing a library containing a large number of potential therapeutic compounds (candidate compounds). Such "combinatorial chemical libraries" are then screened in one or more assays to identify those library members (particular chemical species or subclasses) that display a desired characteristic activity. The compounds thus identified can serve as conventional "lead compounds" or can themselves be used as potential or actual therapeutics.

A combinatorial chemical library is a collection of diverse chemical compounds generated by either chemical synthesis or biological synthesis by combining a number of chemical "building blocks" such as reagents. For example, a linear combinatorial chemical library, such as a polypeptide (e.g., mutein) library, is formed by combining a set of chemical building blocks called amino acids in every possible way for a given compound length (i.e., the number of amino acids in a polypeptide compound). Millions of chemical compounds can be synthesized through such combinatorial mixing of chemical building blocks (Gallop *et al.*, 1994, *J. Med. Chem.* 37(9):1233-1251).

Preparation and screening of combinatorial chemical libraries is well known to those of skill in the art. Such combinatorial chemical libraries include, but are not limited to, peptide libraries, peptoids, encoded peptides, random

bio-oligomers, nonpeptidal peptidomimetics, analogous organic syntheses of small compound libraries, nucleic acid libraries, peptide nucleic acid libraries, antibody libraries, carbohydrate libraries and small organic molecule libraries.

The assays to identify modulators are amenable to high throughput screening. Preferred assays thus detect enhancement or inhibition of ovarian cancer gene transcription, inhibition or enhancement of polypeptide expression, and inhibition or enhancement of polypeptide activity.

High throughput assays for the presence, absence, quantification, or other properties of particular nucleic acids or protein products are well known to those of skill in the art. Similarly, binding assays and reporter gene assays are similarly well known. Thus, e.g., U.S. Patent No. 5,559,410 discloses high throughput screening methods for proteins, U.S. Patent No. 5,585,639 discloses high throughput screening methods for nucleic acid binding (i.e., in arrays), while U.S. Patent Nos. 5,576,220 and 5,541,061 disclose high throughput methods of screening for ligand/antibody binding.

In addition, high throughput screening systems are commercially available *(see, e.g.,* Zymark Corp., Hopkinton, MA; Air Technical Industries, Mentor, OH; Beckman Instruments, Inc. Fullerton, CA; Precision Systems, Inc., Natick, MA, *etc.)*. These systems typically automate entire procedures, including all samisle and reagent pipetting, liquid dispensing, timed incubations, and final readings of the microplate in detectors) appropriate for the assay. These configurable systems provide high throughput and rapid start up as well as a high degree of flexibility and customization. The manufacturers of such systems provide detailed protocols for various high throughput systems. Thus, e.g., Zymark Corp. provides technical bulletins describing screening systems for detecting the modulation of gene transcription, ligand binding, and the like.

In one embodiment, modulators are proteins, often naturally occurring proteins or fragments of naturally occurring proteins. Thus, e.g., cellular extracts containing proteins, or random or directed digests of proteinaceous cellular extracts, may be used. In this way libraries of proteins may be made for screening in the methods of the invention. Particularly preferred in this embodiment are libraries of bacterial, fungal, viral, and mammalian proteins, with the latter being preferred, and human proteins being especially preferred. Particularly useful test compound will be directed to the class of proteins to which the target belongs, e.g., substrates for enzymes or ligands and receptors.

In a preferred embodiment, modulators are peptides of from about 5 to about 30 amino acids, with from about 5 to about 20 amino acids being preferred, and from about 7 to about 15 being particularly preferred. The peptides may be

digests of naturally occurring proteins as is outlined above, random peptides, or "biased" random peptides. By "randomized" or grammatical equivalents herein is meant that each nucleic acid and peptide consists of essentially random nucleotides and amino acids, respectively. Since generally these random peptides (or nucleic acids, discussed below) are chemically synthesized, they may incorporate any nucleotide or amino acid at any position. The synthetic process can be designed to generate randomized proteins or nucleic acids, to allow the formation of all or most of the possible combinations over the length of the sequence, thus forming a library of randomized candidate bioactive proteinaceous agents.

In one embodiment, the library is fully randomized, with no sequence preferences or constants at any position. In a preferred embodiment, the library is biased. That is, some positions within the sequence are either held constant, or are selected from a limited number of possibilities. For example, in a preferred embodiment, the nucleotides or amino acid residues are randomized within a defined class, e.g., of hydrophobic amino acids, hydrophilic residues, sterically biased (either small or large) residues, towards the creation of nucleic acid binding domains, the creation of cysteines, for cross-linking, prolines for SH-3 domains, serines, threonines, tyrosines or histidines for phosphorylation sites, etc., or to purines, etc.

Modulators of ovarian cancer can also be nucleic acids, as defined below. As described above generally for proteins, nucleic acid modulating agents may be naturally occurring nucleic acids, random nucleic acids, or "biased" random nucleic acids. For example, digests of procaryotic or eucaryotic genomes may be used as is outlined above for proteins.

In certain embodiments, the activity of a ovarian cancer-associated protein is down-regulated, or entirely inhibited, by the use of antisense polynucleotide, *i.e.*, a nucleic acid complementary to, and which can preferably hybridize specifically to, a coding mRNA nucleic acid sequence, e.g., a ovarian cancer protein mRNA, or a subsequence thereof. Binding of the antisense polynucleotide to the mRNA reduces the translation and/or stability of the mRNA.

In the context of this invention, antisense polynucleotides can comprise naturally-occurring nucleotides, or synthetic species formed from naturally-occurring subunits or their close homologs. Antisense polynucleotides may also have altered sugar moieties or inter-sugar linkages. Exemplary among these are the phosphorothioate and other sulfur containing species which are known for use in the art. Analogs are comprehended by this invention so long as

they function effectively to hybridize with the ovarian cancer protein mRNA. *See, e.g.*, Isis Pharmaceuticals, Carlsbad, CA; Sequitor, Inc., Natick, MA.

Such antisense polynucleotides can readily be synthesized using recombinant means, or can be synthesized *in vitro*. Equipment for such synthesis is sold by several vendors, including Applied Biosystems. The preparation of other oligonucleotides such as phosphorothioates and alkylated derivatives is also well known to those of skill in the art.

Antisense molecules as used herein include antisense or sense oligonucleotides. Sense oligonucleotides can, e.g., be employed to block transcription by binding to the anti-sense strand. The antisense and sense oligonucleotide comprise a single-stranded nucleic acid sequence (either RNA or DNA) capable of binding to target mRNA (sense) or DNA (antisense) sequences for ovarian cancer molecules. Antisense or sense oligonucleotides, according to the present invention, comprise a fragment generally at least about 14 nucleotides, preferably from about 14 to 30 nucleotides. The ability to derive an antisense or a sense oligonucleotide, based upon a cDNA sequence encoding a given protein is described in, e.g., Stein & Cohen *(Cancer Res.* 48:2659 (1988 and van der Krol *et al. (BioTechniques* 6:958 (1988)).

In addition to antisense polynucleotides, ribozymes can be used to target and inhibit transcription of ovarian cancer-associated nucleotide sequences. A ribozyme is an RNA molecule that catalytically cleaves other RNA molecules. Different kinds of ribozymes have been described, including group I ribozymes, hammerhead ribozymes, hairpin ribozymes, RNase P, and axhead ribozymes *(see, e.g.,* Castanotto *et al., Adv. in Pharmacology 25: 289-317 (1994)* for a general review of the properties of different 5 ribozymes).

Methods of preparing ribozymes are well known to those of skill in the art *(see, e.g., WO 94/26877;* Ojwang *et al., Proc. Natl. Acad. Sci. USA 90:6340-6344 (1993);* Yamada *et al., Human Gene Therapy 1:39-45 (1994);* Leavitt *et al., Proc. Natl. Acad. Sci. USA 92:699- 703 (1995);* Leavitt *et al., Human Gene Therapy 5:1151-120 (1994);* and Yamada *et al., Virology 205: 121-126 (1994)).*

Polynucleotide modulators of ovarian cancer may be introduced into a cell containing the target nucleotide sequence by formation of a conjugate with a ligand binding molecule, as described in WO 91/04753. Suitable ligand binding molecules include, but are not limited to, cell surface receptors, growth factors, other cytokines, or other ligands that bind to cell surface receptors. Preferably, conjugation of the ligand binding molecule does not substantially interfere with the ability of the ligand binding molecule to bind to its corresponding molecule or receptor, or block entry of the sense or antisense oligonucleotide or its

conjugated version into the cell. Alternatively, a polynucleotide modulator of ovarian cancer may be introduced into a cell containing the target nucleic acid sequence, e.g., by formation of an polynucleotide-lipid complex, as described in WO 90/10448. It is understood that the use of antisense molecules or knock out and knock in models may also be used in screening assays as discussed above, in addition to methods of treatment.

As noted above, gene expression monitoring is conveniently used to test candidate modulators (e.g., protein, nucleic acid or small molecule). After the candidate agent has been added and the cells allowed to incubate for some period of time, the sample containing a target sequence to be analyzed is added to the biochip. If required, the target sequence is prepared using known techniques. For example, the sample may be treated to lyse the cells, using known lysis buffers, electroporation, etc., with purification and/or amplification such as PCR performed as appropriate. For example, an *in vitro* transcription with labels covalently attached to the nucleotides is performed. Generally, the nucleic acids are labeled with biotin-FITC or PE, or with cy3 or cy5.

In a preferred embodiment, the target sequence is labeled with, e.g., a fluorescent, a chemiluminescent, a chemical, or a radioactive signal, to provide a means of detecting the target sequence's specific binding to a probe. The label also can be an enzyme, such as, alkaline phosphatase or horseradish peroxidase, which when provided with an appropriate substrate produces a product that can be detected. Alternatively, the label can be a labeled compound or small molecule, such as an enzyme inhibitor, that binds but is not catalyzed or altered by the enzyme. The label also can be a moiety or compound, such as, an epitope tag or biotin which specifically binds to streptavidin. For the example of biotin, the streptavidin is labeled as described above, thereby, providing a detectable signal for the bound target sequence. Unbound labeled streptavidin is typically removed prior to analysis.

As will be appreciated by those in the art, these assays can be direct hybridization assays or can comprise "sandwich assays", which include the use of multiple probes, as is generally outlined in U.S. Patent Nos. 5,681,702, 5,597,909, 5,545,730, 5,594,117, 5,591,584, 5,571,670, 5,580,731, 5,571,670, 5,591,584, 5,624,802, 5,635,352, 5,594,118, 5,359,100, 5,124,246 and 5,681,697, all of which are hereby incorporated by reference. In this embodiment, in general, the target nucleic acid is prepared as outlined above, and then added to the biochip comprising a plurality of nucleic acid probes, under conditions that allow the formation of a hybridization complex.

A variety of hybridization conditions may be used in the present invention, including high, moderate and low stringency conditions as outlined above. The assays are generally run under stringency conditions which allows formation of the label probe hybridization complex only in the presence of target. Stringency can be controlled by altering a step parameter that is a thermodynamic variable, including, but not limited to, temperature, formamide concentration, salt concentration, chaotropic salt concentration pH, organic solvent concentration, etc.

These parameters may also be used to control non-specific binding, as is generally outlined in U.S. Patent No. 5,681,697. Thus it may be desirable to perform certain steps at higher stringency conditions to reduce non-specific binding.

The reactions outlined herein may be accomplished in a variety of ways. Components of the reaction may be added simultaneously, or sequentially, in different orders, with preferred embodiments outlined below. In addition, the reaction may include a variety of other reagents. These include salts, buffers, neutral proteins, e.g. albumin, detergents, *etc.* which may be used to facilitate optimal hybridization and detection, and/or reduce non-specific or background interactions. Reagents that otherwise improve the efficiency of the assay, such as protease inhibitors, nuclease inhibitors, anti-microbial agents, *etc.*, may also be used as appropriate, depending on the sample preparation methods and purity of the target.

The assay data are analyzed to determine the expression levels, and changes in expression levels as between states, of individual genes, forming a gene expression profile.

Screens are performed to identify modulators of the ovarian cancer phenotype. In one embodiment, screening is performed to identify modulators that can induce or suppress a particular expression profile, thus preferably generating the associated phenotype. In another embodiment, *e.g.*, for diagnostic applications, having identified differentially expressed genes important in a particular state, screens can be performed to identify modulators that alter expression of individual genes. In an another embodiment, screening is performed to identify modulators that alter a biological function of the expression product of a differentially expressed gene. Again, having identified the importance of a gene in a particular state, screens are performed to identify agents that bind and/or modulate the biological activity of the gene product.

In addition screens can be done for genes that are induced in response to a candidate agent. After identifying a modulator based upon its ability to suppress

a ovarian cancer expression pattern leading to a normal expression pattern, or to modulate a single ovarian cancer gene expression profile so as to mimic the expression of the gene from normal tissue, a screen as described above can be performed to identify genes that are specifically modulated in response to the agent. Comparing expression profiles between normal tissue and agent treated ovarian cancer tissue reveals genes that are not expressed in normal tissue or ovarian cancer tissue, but are expressed in agent treated tissue. These agent-specific sequences can be identified and used by methods described herein for ovarian cancer genes or proteins. In particular these sequences and the proteins they encode find use in marking or identifying agent treated cells. In addition, antibodies can be raised against the agent induced proteins and used to target novel therapeutics to the treated ovarian cancer tissue sample.

Thus, in one embodiment, a test compound is administered to a population of ovarian cancer cells, that have an associated ovarian cancer expression profile. By "administration" or "contacting" herein is meant that the candidate agent is added to the cells in such a manner as to allow the agent to act upon the cell, whether by uptake and intracellular action, or by action at the cell surface. In some embodiments, nucleic acid encoding a proteinaceous candidate agent (i.e., a peptide) may be put into a viral construct such as an adenoviral or retroviral construct, and added to the cell, such that expression of the peptide agent is accomplished. Regulatable gene administration systems can also be used.

Once the test compound has been administered to the cells, the cells can be washed if desired and are allowed to incubate under preferably physiological conditions for some period of time. The cells are then harvested and a new gene expression profile is generated, as outlined herein.

Thus, e.g., ovarian cancer tissue may be screened for agents that modulate, e.g., induce or suppress the ovarian cancer phenotype. A change in at least one gene, preferably many, of the expression profile indicates that the agent has an effect on ovarian cancer activity. By defining such a signature for the ovarian cancer phenotype, screens for new drugs that alter the phenotype can be devised. With this approach, the drug target need not be known and need not be represented in the original expression screening platform, nor does the level of transcript for the target protein need to change.

In a preferred embodiment, as outlined above, screens may be done on individual genes and gene products (proteins). That is, having identified a particular differentially expressed gene as important in a particular state, screening of modulators of either the expression of the gene or the gene product itself can be done. The gene products of differentially expressed genes are

sometimes referred to herein as "ovarian cancer proteins" or a "ovarian cancer modulatory protein". The ovarian cancer modulatory protein may be a fragment, or alternatively, be the full length protein to the fragment encoded by the nucleic acids referred to in the Tables. Preferably, the ovarian cancer modulatory protein is a fragment. In a preferred embodiment, the ovarian cancer amino acid sequence which is used to determine sequence identity or similarity is encoded by a nucleic acid referred to in the Tables. In another embodiment, the sequences are naturally occurring allelic variants of a protein encoded by a nucleic acid referred to in the Tables. In another embodiment, the sequences are sequence variants as further described herein.

Preferably, the ovarian cancer modulatory protein is a fragment of approximately 14 to 24 amino acids long. More preferably the fragment is a soluble fragment. Preferably, the fragment includes a non-transmembrane region. In a preferred embodiment, the fragment has an N-terminal Cys to aid in solubility. In one embodiment, the C-terminus of the fragment is kept as a free acid and the N-terminus is a free amine to aid in coupling, i.e., to cysteine.

In one embodiment the ovarian cancer proteins are conjugated to an immunogenic agent as discussed herein. In one embodiment the ovarian cancer protein is conjugated to BSA.

Measurements of ovarian cancer polypeptide activity, or of ovarian cancer or the ovarian cancer phenotype can be performed using a variety of assays. For example, the effects of the test compounds upon the function of the ovarian cancer polypeptides can be measured by examining parameters described above. A suitable physiological change that affects activity can be used to assess the influence of a test compound on the polypeptides of this invention. When the functional consequences are determined using intact cells or animals, one can also measure a variety of effects such as, in the case of ovarian cancer associated with tumours, tumour growth, tumour metastasis, neovascularization, hormone release, transcriptional changes to both known and uncharacterized genetic markers (e.g., northern blots), changes in cell metabolism such as cell growth or pH changes, and changes in intracellular second messengers such as cGMP. In tire assays of the invention, mammalian ovarian cancer polypeptide is typically used, e.g., mouse, preferably human.

Assays to identify compounds with modulating activity can be performed *in vitro*. For example, a ovarian cancer polypeptide is first contacted with a potential modulator and incubated for a suitable amount of time, e.g., from 0.5 to 48 hours. In one embodiment, the ovarian cancer polypeptide levels are determined *in vitro* by measuring the level of protein or mRNA. The level of protein is measured

using immunoassays such as western blotting, ELISA and the like with an antibody that selectively binds to the ovarian cancer polypeptide or a fragment thereof. For measurement of mRNA, amplification, e.g., using PCR, LCR, or hybridization assays, e.g., northern hybridization, RNAse protection, dot blotting, are preferred. The level of protein or mRNA is detected using directly or indirectly labeled detection agents, e.g., fluorescently or radioactively labeled nucleic acids, radioactively or enzymatically labeled antibodies, and the like, as described herein.

Alternatively, a reporter gene system can be devised using the ovarian cancer protein promoter operably linked to a reporter gene such as luciferase, green fluorescent protein, CAT, or (beta-gal. The reporter construct is typically transfected into a cell. After treatment with a potential modulator, the amount of reporter gene transcription, translation, or activity is measured according to standard techniques known to those of skill in the art.

In a preferred embodiment, as outlined above, screens may be done on individual genes and gene products (proteins). That is, having identified a particular differentially expressed gene as important in a particular state, screening of modulators of the expression of the gene or the gene product itself can be done. The gene products of differentially expressed genes are sometimes referred to herein as "ovarian cancer proteins." The ovarian cancer protein may be a fragment, or alternatively, be the full length protein to a fragment shown herein.

In one embodiment, screening for modulators of expression of specific genes is performed. Typically, the expression of only one or a few genes are evaluated. In another embodiment, screens are designed to first find compounds that bind to differentially expressed proteins. These compounds are then evaluated for the ability to modulate differentially expressed activity. Moreover, once initial candidate compounds are identified, variants can be further screened to better evaluate structure activity relationships.

In a preferred embodiment, binding assays are done. In general, purified or isolated gene product is used; that is, the gene products of one or more differentially expressed nucleic acids are made. For example, antibodies are generated to the protein gene products, and standard immunoassays are run to determine the amount of protein present. Alternatively, cells comprising the ovarian cancer proteins can be used in the assays.

Thus, in a preferred embodiment, the methods comprise combining a ovarian cancer protein and a candidate compound, and determining the binding of the compound to the ovarian cancer protein. Preferred embodiments utilize the

human ovarian cancer protein, although other mammalian proteins may also be used, e.g. for the development of animal models of human disease. In some embodiments, as outlined herein, variant or derivative ovarian cancer proteins may be used.

Generally, in a preferred embodiment of the methods herein, the ovarian cancer protein or the candidate agent is non-diffusably bound to an insoluble support having isolated sample receiving areas (e.g. a microtiter plate, an array, etc.). The insoluble supports may be made of any composition to which the compositions can be bound, is readily separated from soluble material, and is otherwise compatible with the overall method of screening. The surface of such supports may be solid or porous and of any convenient shape. Examples of suitable insoluble supports include microtiter plates, arrays, membranes and beads. These are typically made of glass, plastic (e.g., polystyrene), polysaccharides, nylon or nitrocellulose, teflon™, etc. microtitre plates and arrays are especially convenient because a large number of assays can be carried out simultaneously, using small amounts of reagents and samples. The particular manner of binding of the composition is not crucial so long as it is compatible with the reagents and overall methods of the invention, maintains the activity of the composition and is nondiffusable. Preferred methods of binding include the use of antibodies (which do not sterically block either the ligand binding site or activation sequence when the protein is bound to the support), direct binding to "sticky" or ionic supports, chemical crosslinking, the synthesis of the protein or agent on the surface, etc. Following binding of the protein or agent, excess unbound material is removed by washing. The sample receiving areas may then be blocked through incubation with bovine serum albumin (BSA), casein or other innocuous protein or other moiety.

In a preferred embodiment, the ovarian cancer protein is bound to the support, and a test compound is added to the assay. Alternatively, the candidate agent is bound to the support and the ovarian cancer protein is added. Novel binding agents include specific antibodies, non-natural binding agents identified in screens of chemical libraries, peptide analogs, etc. Of particular interest are screening assays for agents that have a low toxicity for human cells. A wide variety of assays may be used for this purpose, including labeled in vitro protein-protein binding assays, electrophoretic mobility shift assays, immunoassays for protein binding, functional assays (phosphorylation assays, etc.) and the like.

The determination of the binding of the test modulating compound to the ovarian cancer protein may be done in a number of ways. In a preferred

embodiment, the compound is labeled, and binding determined directly, e.g., by attaching all or a portion of the ovarian cancer protein to a solid support, adding a labeled candidate agent (e.g., a fluorescent label), washing off excess reagent, and determining whether the label is present on the solid support. Various blocking and washing steps may be utilized as appropriate.

In some embodiments, only one of the components is labeled, e.g., the proteins (or proteinaceous candidate compounds) can be labeled. Alternatively, more than one component can be labeled with different labels, e.g., ^{125}I for the proteins and a fluorophor for the compound. Proximity reagents, e.g., quenching or energy transfer reagents are also useful.

In one embodiment, the binding of the test compound is determined by competitive binding assay. The competitor is a binding moiety known to bind to the target molecule (i.e., a ovarian cancer protein), such as an antibody, peptide, binding partner, ligand, etc. Under certain circumstances, there may be competitive binding between the compound and the binding moiety, with the binding moiety displacing the compound. In one embodiment, the test compound is labeled. Either the compound, or the competitor, or both, is added first to the protein for a time sufficient to allow binding, if present. Incubations may be performed at a temperature which facilitates optimal activity, typically between 4 and 40°C. Incubation periods are typically optimized, e.g., to facilitate rapid high throughput screening. Typically between 0.1 and 1 hour will be sufficient. Excess reagent is generally removed or washed away. The second component is then added, and the presence or absence of the labeled component is followed, to indicate binding.

In a preferred embodiment, the competitor is added first, followed by the test compound. Displacement of the competitor is an indication that the test compound is binding to the ovarian cancer protein and thus is capable of binding to, and potentially modulating, the activity of the ovarian cancer protein. In this embodiment, either component can be labeled. Thus, e.g., if the competitor is labeled, the presence of label in the wash solution indicates displacement by the agent. Alternatively, if the test compound is labeled, the presence of the label on the support indicates displacement.

In an alternative embodiment, the test compound is added first, with incubation and washing, followed by the competitor. The absence of binding by the competitor may indicate that the test compound is bound to the ovarian cancer protein with a higher affinity. Thus, if the test compound is labeled, the presence of the label on the support, coupled with a lack of competitor binding,

may indicate that the test compound is capable of binding to the ovarian cancer protein.

In a preferred embodiment, the methods comprise differential screening to identity agents that are capable of modulating the activity of the ovarian cancer proteins. In this embodiment, the methods comprise combining a ovarian cancer protein and a competitor in a first sample. A second sample comprises a test compound, a ovarian cancer protein, and a competitor. The binding of the competitor is determined for both samples, and a change, or difference in binding between the two samples indicates the presence of an agent capable of binding to the ovarian cancer protein and potentially modulating its activity. That is, if the binding of the competitor is different in the second sample relative to the first sample, the agent is capable of binding to the ovarian cancer protein.

Alternatively, differential screening is used to identify drug candidates that bind to the native ovarian cancer protein, but cannot bind to modified ovarian cancer proteins. The structure of the ovarian cancer protein may be modeled, and used in rational drug design to synthesize agents that interact with that site. Drug candidates that affect the activity of a ovarian cancer protein are also identified by screening drugs for the ability to either enhance or reduce the activity of the protein.

Positive controls and negative controls may be used in the assays. Preferably control and test samples are performed in at least triplicate to obtain statistically significant results. Incubation of all samples is for a time sufficient for the binding of the agent to the protein. Following incubation, samples are washed free of non-specifically bound material and the amount of bound, generally labeled agent determined. For example, where a radiolabel is employed, the samples may be counted in a scintillation counter to determine the amount of bound compound.

A variety of other reagents may be included in the screening assays. These include reagents like salts, neutral proteins, e.g. albumin, detergents, etc. which may be used to facilitate optimal protein-protein binding and/or reduce non-specific or background interactions. Also reagents that otherwise improve the efficiency of the assay, such as protease inhibitors, nuclease inhibitors, anti-microbial agents, etc., may be used. The mixture of components may be added in an order that provides for the requisite binding.

In a preferred embodiment, the invention provides methods for screening for a compound capable of modulating the activity of a ovarian cancer protein. The methods comprise adding a test compound, as defined above, to a cell comprising ovarian cancer proteins. Preferred cell types include almost any cell.

The cells contain a recombinant nucleic acid that encodes a ovarian cancer protein. In a preferred embodiment, a library of candidate agents are tested on a plurality of cells.

In one aspect, the assays are evaluated in the presence or absence or previous or subsequent exposure of physiological signals, e.g. hormones, antibodies, peptides, antigens, cytokines, growth factors, action potentials, pharmacological agents including chemotherapeutics, radiation, carcinogenics, or other cells (i.e. cell-cell contacts). In another example, the determinations are determined at different stages of the cell cycle process.

In this way, compounds that modulate ovarian cancer agents are identified. Compounds with pharmacological activity are able to enhance or interfere with the activity of the ovarian cancer protein. Once identified, similar structures are evaluated to identify critical structural feature of the compound.

In one embodiment, a method of inhibiting ovarian cancer cell division is provided. The method comprises administration of a ovarian cancer inhibitor. In another embodiment, a method of inhibiting ovarian cancer is provided. The method comprises administration of a ovarian cancer inhibitor. In a further embodiment, methods of treating cells or individuals with ovarian cancer are provided. The method comprises administration of a ovarian cancer inhibitor.

In one embodiment, a ovarian cancer inhibitor is an antibody as discussed above. In another embodiment, the ovarian cancer inhibitor is an antisense molecule.

A variety of cell growth, proliferation, and metastasis assays are known to those of skill in the art, as described below.

Soft agar growth or colony formation in suspension

Normal cells require a solid substrate to attach and grow. When the cells are transformed, they lose this phenotype and grow detached from the substrate. For example, transformed cells can grow in stirred suspension culture or suspended in semi-solid media, such as semi-solid or soft agar. The transformed cells, when transfected with tumour suppressor genes, regenerate normal phenotype and require a solid substrate to attach and grow. Soft agar growth or colony formation in suspension assays can be used to identify modulators of ovarian cancer sequences, which when expressed in host cells, inhibit abnormal cellular proliferation and transformation. A therapeutic compound would reduce or eliminate the host cells' ability to grow in stirred suspension culture or suspended in semisolid media, such as semi-solid or soft.

Techniques for soft agar growth or colony formation in suspension assays are described in Freshney, *Culture of Animal Cells a Manual of Basic Technique* (3rd ed., 1994), herein incorporated by reference. *See also,* the methods section of Garkavtsev *et al.* (1996), supra, herein incorporated by reference.

Contact inhibition and density limitation of growth

Normal cells typically grow in a flat and organized pattern in a petri dish until they touch other cells. When the cells touch one another, they are contact inhibited and stop growing. When cells are transformed, however, the cells are not contact inhibited and continue to grow to high densities in disorganized foci. Thus, the transformed cells grow to a higher saturation density than normal cells. This can be detected morphologically by the formation of a disoriented monolayer of cells or rounded cells in foci within the regular pattern of normal surrounding cells. Alternatively, labeling index with (^{3}H)-thymidine at saturation density can be used to measure density limitation of growth. *See* Freshney (1994), supra. The transformed cells, when transfected with tumour suppressor genes, regenerate a normal phenotype and become contact inhibited and would grow to a lower density.

In this assay, labeling index with (^{3}H)-thymidine at saturation density is a preferred method of measuring density limitation of growth. Transformed host cells are transfected with a ovarian cancer-associated sequence and are grown for 24 hours at saturation density in non-limiting medium conditions. The percentage of cells labeling with (^{3}H)-thymidine is determined autoradiographically. *See,* Freshney (1994), supra.

Growth factor or serum dependence

Transformed cells have a lower serum dependence than their normal counterparts *(see, e.g.,* Temin, J. *Natl. Cancer Insti. 37:167-175 (1966);* Eagle *et al., J. Exp. Med. 131:836-879 (1970));* Freshney, *supra.* This is in part due to release of various growth factors by the transformed cells. Growth factor or serum dependence of transformed host cells can be compared with that of control. *Tumor specific markers levels* Tumor cells release an increased amount of certain factors (hereinafter "tumour specific markers") than their normal counterparts. For example, plasminogen activator (PA) is released from human glioma at a higher level than from normal brain cells *(see, e.g.,* Gullino, *Angiogenesis, tumour vascularization, and potential interference with tumour growth.* in *Biological Responses in Cancer, pp. 178-184* (Mihich (ed.) *1985)).* Similarly, Tumor angiogenesis factor (TAF) is released at a higher level in tumour

cells than their normal counterparts. *See, e.g.,* Folkman, *Angiogenesis and Cancer, Sem Cancer Biol. (1992)).* Various techniques which measure the release of these factors are described in Freshney *(1994), supra.* Also, *see,* Unkless *et al. , J. Biol. Chem. 249:4295-4305 (1974);* Strickland & Beers, J. *Biol. Chem. 251:5694-5702 (1976);* Whur *et al., Br. J. Cancer 42:305 312 (1980);* Gullino, *Angiogenesis, tumour vascularization, and potential interference with tumour growth.* in *Biological Responses in Cancer, pp. 178-184* (Mihich (ed.) *1985);* Freshney *Anticancer Res. 5:111-130 (1985).*

Invasiveness into Matrigel

The degree of invasiveness into Matrigel-or some other extracellular matrix constituent can be used as an assay to identify compounds that modulate ovarian cancer-associated sequences. Tumor cells exhibit a good correlation between malignancy and invasiveness of cells into Matrigel or some other extracellular matrix constituent. In this assay, tumourigenic cells are typically used as host cells. Expression of a tumour suppressor gene in these host cells would decrease invasiveness of the host cells.

Techniques described in Freshney *(1994), supra,* can be used. Briefly, the level of invasion of host cells can be measured by using filters coated with Matrigel or some other extracellular matrix constituent. Penetration into the gel, or through to the distal side of the filter, is rated as invasiveness, and rated histologically by number of cells and distance moved, or by prelabeling the cells with 125 1 and counting the radioactivity on the distal side of the filter or bottom of the dish. *See, e.g.,* Freshney *(1984), supra.*

Tumor growth in vivo

Effects of ovarian cancer-associated sequences on cell growth can be tested in transgenic or immune-suppressed mice. Knock-out transgenic mice can be made, in which the ovarian cancer gene is disrupted or in which a ovarian cancer gene is inserted. Knock- out transgenic mice can be made by insertion of a marker gene or other heterologous gene into the endogenous ovarian cancer gene site in the mouse genome via homologous recombination. Such mice can also be made by substituting the endogenous ovarian cancer gene with a mutated version of the ovarian cancer gene, or by mutating the endogenous ovarian cancer gene, e.g., by exposure to carcinogens.

A DNA construct is introduced into the nuclei of embryonic stem cells. Cells containing the newly engineered genetic lesion are injected into a host mouse embryo, which is re-implanted into a recipient female. Some of these embryos

develop into chimeric mice that possess germ cells partially derived from the mutant cell line. Therefore, by breeding the chimeric mice it is possible to obtain a new line of mice containing the introduced genetic lesion *(see, e.g.,* Capecchi *et al., Science 244:1288 (1989))*. Chimeric targeted mice can be derived according to Hogan *et al., Manipulating the Mouse Embryo: A Laboratory Manual,* Cold Spring Harbor Laboratory *(1988)* and *Teratocarcinomas and Embryonic Stem Cells: A Practical Approach,* Robertson, ed., IRL Press, Washington, *D.C., (1987).*

Alternatively, various immune-suppressed or immune-deficient host animals can be used. For example, genetically athymic "nude" mouse *(see, e.g.,* Giovanella *et al., J. Natl. Cancer Inst. 52:921 (1974))*, a SCID mouse, a thymectomized mouse, or an irradiated mouse *(see, e.g.,* Bradley *et al., Br. J. Cancer 38:263 (1978);* Selby *et al., Br. J. Cancer 41:52 (1980))* can be used as a host. Transplantable tumour cells (typically about 10^6 cells) injected into isogenic hosts will produce invasive tumours in a high proportions of cases, while normal cells of similar origin will not. In hosts which developed invasive tumours, cells expressing a ovarian cancer-associated sequences are injected subcutaneously. After a suitable length of time, preferably 4 to 8 weeks, tumour growth is measured (e.g. by volume or by its two largest dimensions) and compared to the control. Tumours that have a statistically significant reduction (using, e.g. Student's T test) are said to have inhibited growth.

Administration

Proteins and polynucleotides of the invention and substances identified or identifiable by the assay methods of the invention as modulating the activity of the ovarian cancer-associated proteins/polynucleotides can be administered to patients, therapeutically. Typically, such proteins/polynucleotides and substances may preferably be combined with various components to produce compositions of the invention. Preferably the compositions are combined with a pharmaceutically acceptable carrier or diluent to produce a pharmaceutical composition (which may be for human or animal use). Suitable carriers and diluents include isotonic saline solutions, for example phosphate-buffered saline. The composition of the invention may be administered by direct injection. The composition may be formulated for parenteral, intramuscular, intravenous, subcutaneous, intraocular, oral, vaginal or transdermal administration. Typically, each protein may be administered at a dose of from 0.01 to 30 mg/kg body weight, preferably from 0.1 to 10 mg/kg, more preferably from 0.1 to 1 mg/kg body weight.

Polynucleotides/vectors encoding polypeptide components for use in modulating the activity of the ovarian cancer-associated proteins/polynucleotides may be administered directly as a naked nucleic acid construct. When the polynucleotides/vectors are administered as a naked nucleic acid, the amount of nucleic acid administered may typically be in the range of from 1 μg to 10 mg, preferably from 100 μg to 1 mg.

Uptake of naked nucleic acid constructs by mammalian cells is enhanced by several known transfection techniques for example those including the use of transfection agents. Example of these agents include cationic agents (for example calcium phosphate and DEAE-dextran) and lipofectants (for example lipofectam™ and transfectam™). Typically, nucleic acid constructs are mixed with the transfection agent to produce a composition.

Preferably the polynucleotide or vector of the invention is combined with a pharmaceutically acceptable carrier or diluent to produce a pharmaceutical composition. Suitable carriers and diluents include isotonic saline solutions, for example phosphate-buffered saline. The composition may be formulated for parenteral, intramuscular, intravenous, subcutaneous, oral, intraocular or transdermal administration.

The pharmaceutical compositions can be administered in a range of unit dosage forms depending on the method of administration. For example, unit dosage forms suitable for oral administration include, powder, tablets, pills, capsules and lozenges. Orally administered dosage forms will typically be formulated to protect the active ingredient from digestion and may therefore be complexed with appropriate carrier molecules and/or packaged in an appropriately resistant carrier. Suitable carrier molecules and packaging materials/barrier materials are known in the art.

The compositions of the invention can be administered for therapeutic or prophylatic treatments. In therapeutic applications, compositions are administered to a patient suffering from a disease (e.g. ovarian cancer) in an amount sufficient to cure or at least partially ameliorate the disease and its complications. An amount adequate to accomplish this is defined as a "therapeutically effective dose". An amount of the composition that is capable of preventing or slowing the development of cancer in a patient is referred to as a "prophylactically effective dose".

The routes of administration and dosages described are intended only as a guide since a skilled practitioner will be able to determine readily the optimum route of administration and dosage for any particular patient and condition.

Experimental – Identification of genes associated with ovarian cancer development and outcome using transcript profiles.

A cohort of 66 primary ovarian cancers and borderline tumours sampled at the time of primary laparotomy, and 4 normal ovary samples were collected for the purpose of transcript profiling. Total RNA was extracted from the fresh-frozen specimens with Trizol reagent (Life Technologies, Rockville, MD) and was reverse transcribed using a primer containing oligo(dT) and a T7 promoter sequence. The resulting cDNAs were then in vitro transcribed in the presence of biotinylated nucleotides (Bio-11-CTP and Bio-16-UTP) using the T7 MEGAscript kit (Ambion, Austin, TX). The biotinylated targets were analyzed using a customized Affymetrix oligonucleotide microarray representative of over 90% of the expressed human genome. Data from these experiments have been mined to identify ovarian-specific genes, genes that are overexpressed in ovarian cancer, and genes that cosegregate with disease outcome following primary laporatomy. These analyses have identified a series of genes that encode overexpressed cell surface molecules and are thus therapeutic targets as well as transcription factors, protein tyrosine phosphatases and protein kinases, and genes involved in cell proliferation and cell adhesion. In addition, we have identified >100 genes that are associated with a greater risk of ovarian cancer clinical and biochemical recurrence and overall survival.

The results are presented in the following tables.

Table 1 – genes whose expression is upregulated in ovarian cancer patients vs non-ovarian cancer patients. The statistical significance of the upregulation (P value) is also given.

Table 2 – genes whose expression is down regulated in ovarian cancer patients vs non-ovarian cancer patients. The statistical significance of the upregulation (P value) is also given.

Table 3 – average ratio of expression of specific genes between ovarian cancer patients vs non-ovarian cancer patients. The higher the ratio the more highly over-expressed the gene is in ovarian cancer patients. Negative ratios indicate under expression in ovarian cancer patients.

Table 4 – genes overexpressed/underexpressed in mucinous ovarian cancer patients (with P values).

Table 5 – genes overexpressed/underexpressed in borderline ovarian cancer patients (with P values or ratios).

Table 6 – statistical correlation between overexpression/underexpression and death of patient. The lower the P value the more likely it is that a patient having altered expression of the gene will die from the cancer.

Table 7 – statistical correlation between overexpression/underexpression and clinical recurrence. The lower the P value the more likely it is that a patient having altered expression of the gene will experience recurrence of the disease.

Table 8 – statistical correlation between overexpression/underexpression and bioclinical recurrence (defined as a clinically significant increase in the levels of the known biochemical marker CA125). The lower the P value the more likely it is that a patient having altered expression of the gene will experience biochemical recurrence of the disease.

All publications mentioned in the above specification are herein incorporated by reference. Various modifications and variations of the described methods and system of the invention will be apparent to those skilled in the art without departing from the scope of the invention. Although the invention has been described in connection with specific preferred embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are apparent to those skilled in molecular biology or related fields are intended to be within the scope of the invention.

Dated this fifth day of September 2002

Garvan Institute of Medical Research
Patent Attorneys for the Applicant:

F B RICE & CO

Table 1 – Upregulated in OC

Accession number	UniGene Mapping	Gene symbol and title	Putative Function	P value
AA446644	Hs.692:235	TACSTD1, tumor-associated calcium signal transducer 1	Lymphocyte antigen, plasma membrane, tumor antigen. Member of the GA733 family.C arcinoma-associated antigen expressed on most normal epithelial cells and gastrointestinal carcinomas and functions as a homotypic calcium-independent cell adhesion molecule. The antigen is being used as a target for immunotherapy treatment of human carcinomas.	0
AK001782	Hs.15093:210, Hs.290304:1	HSPC195, hypothetical protein HSPC195	Homo sapiens cDNA FLJ10920 fis, clone OVARC1000384-resourcerer.	0
AW296454	Hs.24743:94	FLJ20171, hypothetical protein FLJ20171	contains 3 RNA recognition motifs	0
AW419196	Hs.257924:13	FLJ13782, Hypothetical protein FLJ13782	weakly similar to a drosophila transcription factor	0
AW630088	Hs.76550:164	Homo sapiens mRNA; cDNA DKFZp564B1264 (from clone DKFZp564B1264)	Mal2 T-cell differentiation protein; found thru interaction with TPD52 which is overexpressed in breast cancer; 4 TM may be involved in vesicle transport	0
BE409838	Hs.194657:233	CDH1, cadherin 1, type 1, E-cadherin (epithelial)	Tumor suppressor. Ca2+-dependent glycoprotein, mediates cell-cell interactions in epithelial cells. Mutations correlated with gastric, breast, colorectal, thyroid and ovarian cancer. Loss of function thought to contribute to progression in cancer by increasing proliferation, invasion, and/or metastasis. The ectodomain of this protein mediates bacterial adhesion to mammalian cells and the cytoplasmic domain is required for internalization.	0
N41720	Hs.172684:89	VAMP8, vesicle-associated membrane protein 8 (endobrevin)	Early endosome, membrane fraction, non-selective vesicle docking, non-selective vesicle transport, protein complex assembly, synaptic vesicle. Member of a family involved in docking or fusion of synaptic vesicles. Associated with the perinuclear vesicular structures of the early endocytic compartment.	0
NM_004415	Hs.349499, Hs.74316:263	DSP, desmoplakin (DPI, DPII)	Cell shape and cell size control, cell-cell adherens junction, epidermal differentiation, intermediate filament, structural constituent of cytoskeleton. Acts as a site of attachment for intermediate filaments in desmosomes (intercellular junction in vertebrate epithelial cells). Compound heterozygosity for non-sense and mis-sense mutations underlies skin fragility/woolly hair syndrome.	0
NM_013230	Hs.286124:357	CD24: CD24 antigen (small cell lung carcinoma cluster 4 antigen)	Plasma membrane, humoral defense mechanism. Cell surface antigen; glycosyl phosphatidylinositol (GPI)-linked glycoprotein that differentiates and activates granulocytes and B lymphocytes.	0
U46418	Hs.233950:84,Hs.182265:2,Hs.7771:1	SPINT1, serine protease inhibitor, Kunitz type 1. Hepatocyte growth factor activator inhibitor.	Extracellular, membrane fraction, serine protease inhibitor. Member of the Kunitz family of serine protease inhibitors. Hepatocyte growth factor activator inhibitor is a potent inhibitor specific for HGF activator and is thought to be involved in regulation of proteolytic activation of HGF in injured tissues.	0
AI333660	Hs.17558:16	Homo sapiens, clone IMAGE:4070464, mRNA, partial cds	Function unknown	0.0001
AB020676	Hs.21543:36	KIAA0869, KIAA0869 protein	Function unknown	0.0002

AI282759	Hs.242463:1	KRT8, keratin 8	Cell structure, Cytoskeletal. May form intermediate filaments; type II keratin, member of a family of structural proteins. Disruption of mechanisms that normally regulate keratin expression in vivo could be related to inflammatory and neoplastic pancreatic disorders (Casanova 1999).	0.0002
AI393742	Hs.199067:46	ERBB3, v-erb-b2 erythroblastic leukemia viral oncogene homolog 3 (avian)	Transmembrane receptor protein tyrosine kinase, epidermal growth factor receptor, integral plasma membrane protein, protein amino acid phosphorylation. Member of the ERBB gene family of receptor tyrosine kinases, elevated levels in certain human mammary tumor cell lines. A receptor for heregulin, capable of mediating HGL-stimulated tyrosine phosphorylation of itself. Epidermal growth factor contains both positive and negative determinants for interaction with ErbB-2/ErbB-3 heterodimers (Stortelers 2002)	0.0002
AW957300	Hs.294142:167	ESTs, Weakly similar to CYL1_HUMAN CYLICIN I [H.sapiens]	Function unknown	0.0002
W70171	Hs.75939:33,Hs.170864:1	UMPK, uridine monophosphate kinase	Catalyzes the phosphorylation of uridine monophosphate to uridine diphosphate. First step in production of pyrimidine nucleoside triphosphates required for RNA and DNA synthesis. An allele of this gene may play a role in mediating nonhumoral immunity to Hemophilus influenzae type B.	0.0003
AA165082	Hs.146388:47, Hs.113919:3	MAP7, microtubule-associated protein 7	Establishment and/or maintenance of cell polarity, microtubule associated protein, microtubule cytoskeleton organization and biogenesis, structural molecule. Predominantly expressed in cells of epithelial origin. Involved in microtubule dynamics and cell polarization and differentiation. Stabilizes microtubules, and may modulate microtubule functions. Studies of the related mouse protein suggest an essential role in microtubule function required for spermatogenesis.	0.0004
AA284679	Hs.25640:264, Hs.5372:2	CLDN3, claudin 3	Integral plasma membrane protein, pathogenesis, tight junction, transmembrane receptor. Member of the claudin family of integral membrane proteins; receptor for Clostridium perfringens enterotoxin;	0.0004
AF017307	Hs.166096:170	ELF3, E74-like factor 3 (ets domain transcription factor, epithelial-specific)	Embryogenesis and morphogenesis, transcription co-activator, transcription factor, transcription from Pol II promoter. ETS domain transcriptional activator; activates expression of epithelial cell specific genes.	0.0004
AW247252	Hs.75514:181	NP, nucleoside phosphorylase	DNA modification, nucleobase nucleoside nucleotide and nucleic acid metabolism, purine-nucleoside phosphorylase. Enzyme purine nucleoside phosphorylase together with adenosine deaminase (ADA) serves a key role in purine catabolism, referred to as the salvage pathway. Mutations in either enzyme result in a severe combined immunodeficiency (SCID).	0.0004
NM_015925	Hs.361379, Hs.95697:59, Hs.93649:1	LISCH7, Liver-specific bHLH-Zip transcription factor	LISCH protein	0.0004
AW088642	Hs.97984:22	SOX17, SRY (sex determining region Y)-box 17	Likely ortholog of mouse SRY-box containing gene 17; alias SOX17	0.0005
AI124756	Hs.5337:191	IDH2, isocitrate dehydrogenase 2 (NADP+), mitochondrial	Carbohydrate metabolism, mitochondrion	0.0006
AU076643	Hs.313:273, Hs.297895:1	SPP1, secreted phosphoprotein 1 (osteopontin, bone sialoprotein I, early T-lymphocyte activation 1)	Osteopontin (bone sialoprotein); bone and blood vessel extracellular matrix protein involved in calcification and atherosclerosis. Increased expression is associated with breast tumor metastasis (Urquidi 2002). Role in HCC, especially in cancer-stromal interactions (Gotoh 2002). Association between levels of a biomarker, osteopontin, and ovarian cancer suggest its clinical usefulness (Kim 2002).	0.0006

BE382756	Hs.169902:319, Hs.275406:1	SLC2A1, Solute carrier family 2 (facilitated glucose transporter), member 1	Glucose transporter, membrane fraction. SLC2A1/GLUT1 - facilitated glucose transporter. Glucose transporter is an integral membrane glycoprotein that is involved in transporting glucose into most cells. 12 TMs. Role in transport of glucose across the blood-brain barrier. Consistent marker of ovarian epithelial malignancy (Kalir 2002). Marker for discriminating hepatocellular carcinoma from other carcinomas (Zimmerman 2002).	0.0006
BE512730	Hs.65114:718, Hs.279437:1	KRT18, keratin 18	Cell shape and cell size control, embryogenesis and morphogenesis, intermediate filament, structural constituent of cytoskeleton. Component of intermediate filaments; type I epidermal keratin, strongly similar to murine Endo B. Expressed in single layer epithelial tissues of the body. Mutations linked to cryptogenic cirrhosis.	0.0006
X60111	Hs.1244:227, Hs.230559:1, Hs.242020:1	CD9: CD9 antigen (p24)	Plasma membrane, integral plasma membrane protein. Member of the transmembrane 4 superfamily (TM4SF); may mediate platelet activation and aggregation. Cell surface glycoprotein that is known to complex with integrins and other transmembrane 4 superfamily proteins.	0.0006
AI791905	Hs.95549:147, Hs.229556:1	FLJ20273, RNA-binding protein	Contains four RNA recognition motifs (RRM, RBD, or RNP)	0.0007
NM_006103	Hs.2719:108, Hs.54451:1	WFDC2, WAP four-disulfide core domain 2	Endopeptidase inhibitor, extracellular space, proteolysis and peptidolysis, spermatogenesis. Epididymis-specific secreted protein; may have a role in sperm maturation; may belong to a family of extracellular proteinase inhibitors. Expressed in pulmonary epithelial cells, and also expressed in some ovarian cancers.	0.0009
U81961	Hs.2794:71	SCNN1A, sodium channel, nonvoltage-gated 1 alpha	Amiloride-sensitive sodium channel, excretion, integral plasma membrane protein, membrane fraction, sodium transport. Alpha subunit of the amiloride-sensitive epithelial sodium channel; functions in nonvoltage-gated channel	0.0009
X69699	Hs.73149:72, Hs.213008:1	PAX8, paired box gene 8	Histogenesis and organogenesis, embryogenesis and morphogenesis, thyroid-stimulating hormone receptor, transcription factor. Member of the paired domain family of nuclear transcription factors; may be involved in the ribosome assembly, required for normal thyroid development. PAX genes play critical roles during fetal development and cancer growth.	0.0009
AI027643	Hs.120912:12	ESTs	Function unknown	0.001
AA173992	Hs.7956:28	ESTs	Function unknown	0.0011
AB018249	Hs.10458:10	SCYA16, small inducible cytokine subfamily A (Cys-Cys), member 16.	Antimicrobial humoral response (sensu Invertebrata), cell-cell signaling, chemokine chemotaxis. Cytokine A16; lymphocyte and monocyte chemoattractant.	0.0011
NM_014791	Hs.184339:27	MELK, likely ortholog of maternal embryonic leucine zipper kinase.	KIAA0175 gene product; serine/threonine protein kinase domain	0.0011
AI392640	Hs.18272:81	SLC38A1, solute carrier family 38, member 1	amino acid transporter A1 (ATA1), likely ortholog of mouse N-system amino acid transporter protein NAT2.	0.0012
AW837046	Hs.6527:201	GPR56, G protein-coupled receptor 56	cell adhesion, cell-cell signalling, G-protein linked receptor, integral plasma membrane protein, G-protein linked receptor protein signalling pathway. Member of the G protein-coupled receptor family; similar to secretin and calcitonin receptors. 7 transmembrane domains, a mucin-like domain and cysteine box in the N-terminal region. Expressed in range of tissues, highest levels in thyroid, selectively within the monolayer of cuboidal epithelial cells of the smaller, more actively secreting follicles of human thyroid. Differentially expressed in melanoma cell lines with different metastatic potential (Zendman et al 1999).	0.0012

AI669760	Hs.188881:6, Hs.199354:1	ESTs	dbEST Library Tissue Type restricted to prostate	0.0013
AF132818	Hs.84728:127	KLF5, Kruppel-like factor 5 (intestinal)	RNA polymerase II transcription factor, transcription from Pol II promoter. Zinc finger transcriptional activator; localizes to the nucleus and binds the epidermal growth factor response element, binds GC boxes.	0.0014
AI355761	Hs.242463:2	KRT8, keratin 8	Cell structure, Cytoskeletal. May form intermediate filaments; type II keratin, member of a family of structural proteins. Disruption of mechanisms that normally regulate keratin expression in vivo could be related to inflammatory and neoplastic pancreatic disorders (Casanova 1999).	0.0014
BE019020	Hs.85838:171	SLC16A3, solute carrier family 16 (monocarboxylic acid transporters), member 3 (MCT3)	Integral plasma membrane protein, membrane fraction, monocarboxylic acid transport, monocarboxylic acid transporter. Member of monocarboxylate transporter family; may function as a transporter (MCT3).	0.0015
AA340864	Hs.278562:101	CLDN7, claudin 7	Integral membrane protein, tight junction. Similar to murine Cldn7;	0.0016
AL039104	Hs.159557:394	KPNA2, karyopherin alpha 2 (RAG cohort 1, importin alpha 1)	DNA metabolism, G2 phase of mitotic cell cycle. NLS-bearing substrate-nucleus import, cytoplasm, importin alpha-subunit, nuclear localization sequence binding, nucleoplasm, regulation of DNA recombination, spindle pole body and microtubule cycle (sensu Saccharomyces). Karyopherin alpha 2 (importin alpha 1); subunit of the NLS (nuclear localization signal) receptor. KPNA2 protein interacts with the NLSs of DNA helicase Q1 and SV40 T antigen and may be involved in the nuclear transport of proteins. KPNA2 also may play a role in V(D)J recombination.	0.0016
AW176120	Hs.9061:77	MGC2477, hypothetical protein MGC2477	function unknown	0.0016
BE265489	Hs.3123:49	LLGL2, lethal giant larvae (Drosophila) homolog 2	Cytoskeleton, structural molecule. May associate with nonmuscle myosin II heavy chain. cDNA source cancer cell lines. 57% ID to m.musculus 1920362A tumor suppressor gene mgl1	0.0016
BE279383	Hs.26557:77	PKP3, plakophilin 3	Cell adhesion, intercellular junction. Desmosomal plaque proteins are members of the 'armadillo-repeat' multigene family and have important functions in cytoskeleton/cell membrane interactions.	0.0016
J05581	Hs.89603:128, Hs.296789:1	MUC1, mucin 1, transmembrane	Integral plasma membrane protein. Cell surface mucin glycoprotein expressed by most glandular and ductal epithelial cells and some hematopoietic cell lineages. Alterations in glycosylation in epithelial cancer cells. Marker for hepatocellular carcinoma. MUC1 metabolic complex conserved in tumor-derived and normal epithelial cells. Expression predictor of surgical outcome in mass-forming intrahepatic cholangiocarcinoma. Tyrosine kinase c-Src constitutes a bridge between cystic fibrosis transmembrane regulator channel failure and MUC1 overexpression in cystic fibrosis.	0.0016
AA531276	Hs.59509:9	ESTs (unnamed protein product)	Function unknown	0.0017
AW167128	Hs.231934:3	ESTs; weakly similar to A57717 transcription factor EC2	Function unknown	0.0018
AW368226	Hs.67928:25, Hs.229840:1	Ets-related transcription factor, ESX, epithelium-restricted Ets protein ESX-not in Unigene, but found using resourcerer.	Embryogenesis and morphogenesis, transcription co-activator, transcription factor, transcription from Pol II promoter.	0.0021
AK000733	Hs.23900:82	RACGAP1, Rac GTPase activating protein 1	Strongly similar to murine Racgap1 GTPase-activating protein for rac. The plexin-B1/Rac interaction inhibits PAK activation and enhances Sema4D ligand binding	0.0024
H60720	Hs.81892:95	KIAA0101 gene product	function unknown; no signficant hits with Superfamily	0.0025

NM_014586	Hs.109437:17	HUNK, hormonally upregulated neu tumor-associated kinase	Developmental processes, protein serine/threonine kinase, signal transduction, protein kinase containing SNF1 (fam of serine/threonine kinases) domain; progesterone and estradiol regulated. Similar to murine Hunk.	0.0025
AI885516	Hs.95612:31, Hs.251688:1	desmocollin type 2a, desmocollin 2, isoform Dsc2b preproprotein; desmosomal glycoprotein II/III; desmocollin-3-not in Unigene, but found using resourcerer.	Cell adhesion, intercellular junction	0.0027
AW194426	Hs.20726:17	ESTs	Function unknown	0.0027
NM_001982	Hs.199067:83, Hs.167386:1	ERBB3, HER3 (c-erb-B3), v-erb-b2 erythroblastic leukemia viral oncogene homolog 3 (avian)	Epidermal growth factor receptor, integral plasma membrane protein, protein amino acid phosphorylation. Member of the ERBB gene family of receptor tyrosine kinases, elevated levels in certain human mammary tumor cell lines. A receptor for heregulin, capable of mediating HGL-stimulated tyrosine phosphorylation of itself.	0.0028
NM_007019	Hs.93002:85	UBE2C, ubiquitin carrier protein E2-C	Ubiquitin-dependent protein degradation, degradation of cyclin, protein modification, positive control of cell proliferation. Subunit of a complex with ubiquitin ligase activity; complex that exhibits cyclin-selective ubiquitin ligase activity.	0.0031
BE184455	Hs.251754:128, Hs.245742:1	SLPI, secretory leukocyte protease inhibitor (antileukoproteinase)	Plasma protein, proteinase inhibitor. Secreted inhibitor which protects epithelial tissues from serine proteases. Found in various secretions including seminal plasma, cervical mucus, and bronchial secretions; has affinity for trypsin, leukocyte elastase, and cathepsin G. Its inhibitory effect contributes to the immune response by protecting epithelial surfaces from attack by endogenous proteolytic enzymes; the protein is also thought to have broad-spectrum anti-biotic activity.	0.0034
Y00815	Hs.75216:262, Hs.228792:1, Hs.245063:1	PTPRF, protein tyrosine phosphatase, receptor type, F	Cell adhesion, integral plasma membrane protein, transmembrane receptor protein, tyrosine phosphatase signaling pathway. Receptor-type protein tyrosine phosphatase F; interacts with the insulin receptor; has Ig-like and FN-III repeats in the extracellular domain	0.0035
AA706017	Hs.119944:14	ESTs	Function unknown	0.0038
AA256641	Hs.236894:24	ESTs, Highly similar to S02392 alpha-2-macroglobulin receptor precursor	Function unknown	0.0041
AW055308	Hs.31803:15	ESTs, Weakly similar to TRHY_HUMAN TRICHOHYALI [H.sapiens]	Function unknown	0.0043
AI301558	Hs.290801:35, Hs.356228	EST	Function unknown	0.0044
T18997	Hs.180372:119	BCL2-like 1, Homo sapiens cDNA FLJ20750 fis, clone HEP05174 (hypothetical protein (Macaca fascicularis))	Function unknown	0.0044
AI798863	Hs.87191:8	ESTs	Function unknown	0.0049
J03258	Hs.2062:146	VDR, vitamin D (1,25- dihydroxyvitamin D3) receptor	DNA binding, signal transduction, vitamin D3 receptor. Zinc-finger DNA-binding transcription factor. Genetic polymorphism determines bone mineral density. Stat1-vitamin D receptor interactions antagonize 1,25-dihydroxyvitamin D transcriptional activity and enhance stat1-mediated transcription.	0.0049

AA151647	Hs.68877:141, Hs.228686:1	CYBA, cytochrome b-245, alpha polypeptide	Cytochrome b, membrane, mitochondrion, superoxide metabolism. Alpha-subunit of cytochrome b245, primary component of the microbicidal oxidase system of phagocytes. CYBA deficiency is associated with chronic granulomatous disease (CGD).	0.005
AI538613	Hs.135657:8	TMPRSS3 Transmembrane protease, serine 3	Integral membrane protein, proteolysis and peptidolysis. Contains a serine protease domain, a transmembrane domain, a LDL receptor-like domain, and a scavenger receptor cysteine-rich domain. Serine proteases are known to be involved in a variety of biological processes, whose malfunction often leads to human diseases and disorders. Expressed in fetal cochlea and many other tissues, and is thought to be involved in the development and maintenance of the inner ear or the contents of the perilymph and endolymph. Missense mutations in autosomal recessive sensorineural deafness. Identified as a tumor associated gene that is overexpressed in ovarian tumors.	0.0051
AK000978	Hs.79741:18	FLJ10116, hypothetical protein FLJ10116	Function unknown	0.0051
AV662037	Hs.124740:18	Homo sapiens cDNA FLJ30532 fis, clone BRAWH2001129, weakly similar to OCCLUDIN	59% identity to human Zinc finger protein 91	0.0051
AJ278016	Hs.55565:35	ANKRD3, ankyrin repeat domain 3	ATP binding, protein amino acid phosphorylation, protein binding, protein serine/threonine kinase.	0.0055
AL046341	Hs.75562:147	DDR1, discoidin domain receptor family, member 1	Cell adhesion, integral plasma membrane protein, transmembrane receptor, protein tyrosine kinase. Epithelial-specific receptor protein tyrosine kinase; may be involved in cell adhesion; has putative discoidin motifs in extracellular domain. DDR1 (CD167a) is a RTK that is widely expressed in normal and transformed epithelial cells and is activated by various types of collagen.	0.0055
T09997	Hs.70327:196, Hs.211478:1	CRIP-2, cysteine-rich protein 2	Zn-finger LIM domain protein;208-amino acid protein containing 2 LIM domains	0.0055
BE302796	Hs.105097:115	TK1, thymidine kinase 1, soluble	Cytoplasm, thymidine kinase. Generates thymidylate for DNA synthesis. TK1 gene expression together with TS, TP and DPD gene expression may play important roles in influencing the malignant behavior of epithelial ovarian cancer (Fujiwaki R 2002).	0.006
J04088	Hs.156346:184, Hs.270810:2	TOP2A, topoisomerase (DNA) II alpha (170kD)	DNA binding, DNA topoisomerase (ATP-hydrolyzing), nucleus. DNA topoisomerase II alpha; may relax DNA torsion upon replication or transcription. Involved in processes such as chromosome condensation, chromatid separation, and the relief of torsional stress that occurs during DNA transcription and replication. Catalyzes the transient breaking and rejoining of two strands of duplex DNA. The gene encoding this enzyme functions as the target for several anticancer agents and a variety of mutations in this gene have been associated with the development of drug resistance. Reduced activity of this enzyme may also play a role in ataxia-telangiectasia.	0.006
U46455	Hs.252189:148, Hs.248217:1	SDC4, syndecan 4 (amphiglycan, ryudocan)	Integral plasma membrane, proteoglycan syndecan. Syndecans are transmembrane heparan sulfate proteoglycans that appear to act as receptors or coreceptors involved in intracellular communication. Members of the MYC gene family and 4 members of the syndecan gene family are closely situated on 4 different chromosomes.	0.0061
M79141	Hs.13234:39	ESTs	Function unknown	0.0062
AI955040	Hs.301584:5, Hs.265398:3	ESTs, Moderately similar to hypothetical protein FLJ20378 [Homo sapiens] [H.sapiens]	Function unknown	0.0065

Z95636	Hs.11669:81, Hs.231010:1	LAMA5, laminin, alpha 5	Basement lamina, structural molecule. Widely expressed in adult tissues, with highest levels in lung, heart, and kidney. Fifth member of the alpha subfamily of vertebrate laminin chains. Possible basement membrane protein; contains laminin EGF-like domain, two extracellular laminin G domains.	0.0066
BE563085	Hs.833:97	ISG15, interferon-stimulated protein, 15 kDa	Cell-cell signaling, cytoplasm, extracellular space, protein binding. Protein that is induced by interferon.	0.0068
BE278288	Hs.155048:119	LU, Lutheran blood group (Auberger b antigen included)	Blood group antigen, cell adhesion, integral plasma membrane protein, signal transduction, transmembrane receptor. Lutheran blood group glycoprotein; may play role in cell-cell, cell-matrix adhesion, signal transduction; member of the Ig superfamily, has integrin-binding motifs, SH3 domains.	0.0069
AB040914	Hs.278628:52	ShrmL, Shroom-related protein (KIAA1481 protein)	Amiloride-sensitive sodium channel (weakly similar to Mus musculus PDZ domain actin binding protein)	0.0074
AI262789	Hs.93659:52	ERP70, protein disulfide isomerase related protein (calcium-binding protein, intestinal-related)	Endoplasmic reticulum lumen, protein secretion. Strongly similar to rat Rn.4070 (CABP2); may bind calcium.	0.008
NM_006147	Hs.11801:77	IRF6, interferon regulatory factor 6	Member 6 of the interferon regulatory factor transcription factor family; has low similarity to IRF4, which is a lymphocytic transcription factor that stimulates B cell proliferation.	0.0082
R61463	Hs.16165:50	LAK-4P, expressed in activated T/LAK lymphocytes	expressed in activated T/LAK lymphocytes	0.0082
AI878857	Hs.109706:285	HN1, hematological and neurological expressed 1 protein	Strongly similar to murine Hn1	0.0087
AK001763	Hs.73239:37	FLJ10901, hypothetical protein FLJ10901	B link shows some homology to KIAA1294 but no known function	0.009
AC004770	Hs.4756:99	FEN1, flap structure-specific endonuclease 1	DNA repair enzyme, DNA replication, UV protection, double-strand break repair, double-stranded DNA binding, double-stranded DNA specific exodeoxyribonuclease, endonuclease, fatty acid desaturation, membrane fraction. Removes 5' overhanging flaps in DNA repair and processes the 5' ends of Okazaki fragments in lagging strand DNA synthesis.	0.0093
AI567421	Hs.273330:137	AGRN: agrin	Agrin is a neuronal aggregating factor that induces the aggregation of acetylcholine receptors and other postsynaptic proteins on muscle fibers and is crucial for the formation of the neuromuscular junction. Acts at the nerve-muscle synapse in the glomerular basal membrane and on T-lymphocytes.	0.0093
AW161386	Hs.13561:49	MGC4692: hypothetical protein MGC4692	Function unknown	0.0103
M85430	Hs.155191:546	VIL2, villin 2 (ezrin)	Cytoskeletal anchoring, microvillus. Regulates cell adhesion and cortical morphogenesis. The cytoplasmic peripheral membrane protein encoded by this gene functions as a protein-tyrosine kinase substrate in microvilli. As a member of the ERM protein family, this protein serves as an intermediate between the plasma membrane and the actin cytoskeleton. It plays a key role in cell surface structure adhesion, migration, and organization.	0.0106
AW250380	Hs.109059:124, Hs.24756:11	MRPL12, mitochondrial ribosomal protein L12	Protein synthesis, General cellular role, Ribosomal subunit, Mitochondrial, RNA-binding protein, Ribosome-associated.	0.0114
AI733848	Hs.71935:13	ZNF339, zinc finger protein 339	Zinc finger protein	0.0115
AF111856	Hs.105039:48	SLC34A2, solute carrier family 34 (sodium	SLC34A2: solute carrier family 34 (sodium phosphate), member 2; contains 8 predicted TMs and a cysteine-rich N-terminal region. Type 2 sodium-dependent phosphate transporter. member of the renal	0.0121

		phosphate), member 2	type II co-transporter family.	
BE386983	Hs.343214	LOC112616, Hypothetical protein BC010116	Function unknown.	0.0131
AA433988	Hs.98502:8	MUC16, mucin 16, CA125	Mucin 16. Alias CA125 ovarian cancer antigen	0.0137
AW248314	Hs.9622:83	MRPS18A, mitochondrial ribosomal protein S18A	Mitochondrial small ribosomal subunit, protein biosynthesis, structural constituent of ribosomeribosomal mitochondrial protein S18A	0.0149
AA454501	Hs.43666:65	PTP4A3, protein tyrosine phosphatase type IVA, member 3	Prenylated protein tyrosine phosphatase. PTPs are cell signaling molecules that play regulatory roles in a variety of cellular processes. Strong similarity to murine Ptp4a3 (Mm.4124). Overexpression of this gene in mammalian cells was reported to inhibit angiotensin-II induced cell calcium mobilization and promote cell growth. PRL3 (PTP4A3) expressed at high levels cancer metastases (Saha et al. 2001), PRL3 gene is important for colorectal cancer metastasis.	0.016
U33446	Hs.75799:116	PRSS8, protease, serine, 8 (prostasin)	Extracellular space, plasma membrane, serine type peptidase. A trypsinogen, member of the trypsin family of serine proteases. Highly expressed in prostate epithelia, one of several proteolytic enzymes found in seminal fluid. Protease-mediated regulation of sodium absorption is a function of human airway epithelia, and prostasin is a likely candidate for this activity.	0.0166
X98654	Hs.93837:43	PITPNM, phosphatidylinositol transfer protein, membrane-associated	Brain development, lipid metabolism, membrane fraction, phosphatidylinositol transporter, phototransduction. Catalyzes the transfer of phosphatidylinositol between membranes; similar to Drosophila rdgB.	0.0167
AI660149	Hs.44865:39, Hs.300819:19, Hs.293904:14	LEF1, Lymphoid enhancer-binding factor-1	Very strongly similar to murine Lef1; may act as a transcription factor. Expressed in pre-B and T cells. Binds to T-cell receptor-alpha enhancer and confers maximal enhancer activity. A target gene ectopically activated in colon cancer, from selective activation of a promoter for a full-length LEF1 isoform that binds beta-catenin (HOVANES 2001).	0.0172
AF098158	Hs.9329:152	C20orf1, chromosome 20 open reading frame 1	ATP binding, GTP binding, cell proliferation, mitosis, nucleus spindle. Proliferation-associated nuclear protein; associates with the spindle pole and mitotic spindle during mitosis	0.0183
AB014551	Hs.155120:101, Hs.337774	ARHGEF2, rho/rac guanine nucleotide exchange factor (GEF) 2	Cell shape and cell size control, cell surface receptor linked signal transduction, guanyl-nucleotide exchange factor, microtubule cytoskeleton. Rho GTPases play a fundamental role in numerous cellular processes that are initiated by extracellular stimuli that work through G protein coupled receptors. The encoded protein may form complex with G proteins and stimulate Rho-dependent signals. Rho/Rac guanine nucleotide exchange factor (GEF) 2; associates with microtubules, stimulates GTP binding on Rac and Rho	0.0206
AI278023	Hs.89986:24, Hs.290780:1	ESTs	Function unknown	0.0208
Z95152	Hs.178695:25, Hs.79107:1	MAPK13, mitogen-activated protein kinase 13	MAP kinase, antimicrobial humoral response (sensu Invertebrata), cell surface receptor, signal transduction, chemotaxis, stress response. MAP kinases act as an integration point for multiple biochemical signals, and are involved in a wide variety of cellular processes such as proliferation, differentiation, transcription regulation and development. Can be activated by proinflammatory cytokines and cellular stress. Transcription factor ATF2, and microtubule dynamics regulator stathmin are substrates of this kinase.	0.0217

AW840171	Hs.265398:7	ESTs, Moderately similar to hypothetical protein FLJ20378 [Homo sapiens] [H.sapiens]	Function unknown	0.0222
D49441	Hs.155981:53	MSLN, mesothelin	Cell adhesion, cell surface antigen, membrane. Pre-pro-megakaryocyte potentiating factor. An antibody that reacts with ovarian cancers and mesotheliomas was used to isolate a cell surface antigen named mesothelin. Although the function of mesothelin is unknown, it may play a role in cellular adhesion and is present on mesothelium, mesotheliomas, and ovarian cancers.	0.0225
AW797437	Hs.69771:262, Hs.444:1, Hs.294163:1	EST, CM1-UM0039-030400-173-a09	Function unknown	0.0229
BE396290	Hs.5097:261	SYNGR2, synaptogyrin 2	Integral plasma membrane protein, member of a family of transmembrane synaptic vesicle proteins, specialized secretory organelles that store neurotransmitters in nerve terminals, and release them by fusing with the presynaptic plasma membrane during exocytosis.	0.0229
NM_002145	Hs.2733:25	HOXB2, homeo box B2, Hox2H protein	Circulation, developmental processes, transcription factor. Member of homeodomain family of DNA binding proteins; may regulate gene expression, morphogenesis, and differentiation. Genes of the HOXB (or HOX2) complex are expressed specifically in erythromegakaryocytic cell lines, some are expressed only in hematopoietic progenitors.	0.024
AW959311	Hs.87019:8	Hypothetical protein DKFZp434J037	probable serine/threonine protein kinase; KIAA0537	0.0251
BE387202	Hs.118638:166, Hs.276104:1, Hs.276127:1, Hs.276246:1	NME1, non-metastatic cells 1, protein (NM23A)	Transcription factor and nucleoside diphosphate kinase; has a role in the transcriptional regulation of c-myc expression. Mutations in NME1 have been identified in aggressive neuroblastomas.	0.0257
AA379597	Hs.5199:87, Hs.277192:1	HSPC150, HSPC150 protein similar to ubiquitin-conjugating enzyme	Similar to ubiquitin conjugating enzyme	0.0259
BE148235	Hs.193063:100	Homo sapiens cDNA FLJ14201 fis, clone NT2RP3002955	high homology to ARP-3 actin-like protein	0.0259
AF111713	Hs.286218:64	JAM1, junctional adhesion molecule	Cell motility, inflammatory response, intercellular junction. Role in the regulation of tight junction assembly in epithelia. Ligation of JAM is required for reovirus-induced activation of NF-kappa-B and apoptosis. Role in lymphocyte homing.	0.0261
BE391635	Hs.75725:450, Hs.274751:1, Hs.277482:1, Hs.277468:1	TAGLN2, transgelin 2	Complex assembly protein. Homolog of the protein transgelin, which is one of the earliest markers of differentiated smooth muscle. Function not yet determined. May be an actin-binding protein.	0.0275
D14697	Hs.77393:201, Hs.247769:1	FDPS, farnesyl diphosphate synthase (farnesyl pyrophosphate synthetase, dimethylallyltranstransferase, geranyltranstransferase)	Farnesyl pyrophosphate synthetase (farnesyl diphosphate synthase); part of the cholesterol synthesis pathway.	0.0276
AW194364	Hs.94814	MGC2865, Hypothetical protein MGC2865	Function unknown.	0.0295

T47364	Hs.278613:145	IFI27, interferon, alpha-inducible protein 27	Integral membrane protein. Isolated from estradiol-treated human breast carcinoma cells. Induced by interferon-alpha in human cell lines of different origin, expression is independent of the presence of estradiol receptor in the cells.	0.03
U17760	Hs.301103:71, Hs.75517:24, Hs.199068:1	LAMB3, Laminin, beta 3 (nicein (125kD), kalinin (140kD), BM600 (125kD)) (Accn NM_000228)	Epidermal differentiation, laminin-5, structural molecule. Member of a family of basement membrane proteins. LAMB3 serves as the beta chain in laminin-5. Mutations in LAMB3 have been identified as the cause of various types of epidermolysis bullosa.	0.0304
AU076517	Hs.184276:142	SLC9A3R1, solute carrier family 9 (sodium/hydrogen exchanger), isoform 3 regulatory factor 1	Actin cytoskeleton, protein complex assembly. Regulatory cofactor of the NHE3 (SLC9A3) sodium/hydrogen antiporter; interacts with merlin (NF2) and ERM family members; has two PDZ domains. Structural determinants in interaction of beta 2 adrenergic and platelet-derived growth factor receptors	0.0312
AW880841	Hs.96908 , Hs.74427:112	PIG11 , p53-induced protein	Negative control of cell proliferation, stress response. May generate or respond to oxidative stress, may have a role in p53-dependent apoptosis Polyak K, Xia Y, Zweier JL, Kinzler KW, Vogelstein B. A model for p53-induced apoptosis.Nature. 1997 Sep 18;389(6648):300-5.	0.0314
H24185	Hs.92918:91	BM-009, hypothetical protein BM-009	Function unknown	0.0314
BE614410	Hs.23044:51	MGC16386, hypothetical protein, similar to RIKEN cDNA	Function unknown.	0.0326
H16423	Hs.82685:37	CD47: CD47 antigen (Rh-related antigen, integrin-associated signal transducer)	Oncogenesis, plasma membrane, plasma glycoprotein, cell-cell matrix adhesion, integral plasma membrane proteoglycan, integrin receptor signal signalling pathway. Similar to Rh-antigen; may interact with integrins and have a role in intracellular calcium increase during cell adhesion.	0.0336
AU076611	Hs.154672:123	MTHFD2, methylene tetrahydrofolate dehydrogenase (NAD+ dependent); methenyltetrahydrofolate cyclohydrolase	Electron transporter, methenyltetrahydrofolate cyclohydrolase, mitochondrion. encodes a nuclear-encoded mitochondrial bifunctional enzyme with methylenetetrahydrofolate dehydrogenase and methenyltetrahydrofolate cyclohydrolase activities. may provide formyltetrahydrofolate for formylmethionyl tRNA synthesis; involved in initiation of mitochondrial protein synthesis.	0.0342
AI859390	Hs.288940:49	TMEM8, five-span transmembrane protein M83; type I transmembrane protein	Integral plasma membrane protein. Type I transmembrane protein; contains five membrane-spanning domains	0.0345
AA159216	Hs.55505:57	FLJ20442, hypothetical protein FLJ20442	Contains a dual specificity protein phosphatase catalytic domain; 34% similar to protein-tyrosine phosphatase	0.0354
AF119665	Hs.184011:156	PP, pyrophosphatase (inorganic)	Inorganic diphosphatase, phosphate metabolism. Catalyzes the hydrolysis of pyrophosphate to inorganic phosphate	0.0358
BE513613	Hs.11538:275	ARPC1B, actin related protein 2/3 complex, subunit 1A (41 kD)	Cell motility, structural constituent of cytoskeleton. Arp2/3 complex, subunit 1A; involved in assembly of the actin cytoskeleton, may have a role in protrusion of lamellipodia	0.0367
AW992356	Hs.182339	EHF: ets homologous factor	DNA binding, tumor suppressor, cell proliferation, developmental processes, transcription activating factor. Member of the ESE subfamily of Ets transcription factors	0.0404
AW772298	Hs.21103:40, Hs.266784:2, Hs.102950:1	Homo sapiens mRNA; cDNA DKFZp564B076 (from clone DKFZp564B076)	Alias coat protein gamma-cop	0.0423
H16646	Hs.118666:66	PP591, hypothetical protein PP591	Hypothetical protein PP591 (Novel Human cDNA clones with function of inhibiting cancer cell growth; unpublished)	0.043

AA279661	Hs.83753:244, Hs.301236:3	SNRPB, small nuclear ribonucleoprotein polypeptides B and B1	Spliceosome, mRNA splicing, small nuclear ribonucleoprotein. U1 and U2 snRNP protein; component of snRNP complexes, required units of the spliceosome	0.0446
BE001596	Hs.85266:102	ITGB4, integrin, beta 4	Cell adhesion receptor, integrin, invasive growth, oncogenesis. Beta 4 subunit of integrin; involved in cell-cell and cell-matrix interactions; member of a family of cell-surface proteins. Binding of beta 4 to plectin is essential for the proper formation and function of hemidesmosomes.	0.0453
BE246444	Hs.283685:148, Hs.232028:2	FLJ20396, hypothetical protein FLJ20396	100%/175aa unnamed protein g7020468	0.0453
X54942	Hs.83758:34	CKS2, CDC28 protein kinase 2	Cell proliferation, regulation of CDK activity. Similar to S. pombe p13suc1; binds and regulates CDK-cyclin complexes. expressed in different patterns through the cell cycle in HeLa cells, which reflects specialized role for the encoded protein.	0.0478
AA305599	Hs.238205:36	PRO2013, hypothetical protein PRO2013	Function unknown	0.0483
AF019226	Hs.8036:84	RAB3D, member RAS oncogene family	RAB small monomeric GTPase, hemocyte development. GTP-binding protein; may be involved in vesicle transport; member of the RAB family of small GTPases. Alias GOV, that is overexpressed in glioblastoma multiforme tissue as compared to normal brain tissue. GOV is also highly expressed in recurrent glioma, colon tumor metastatic to brain, breast tumors, prostate tumors, and several tumor cell lines	0.0485
NM_001949	Hs.1189:65, Hs.296939:2	E2F3, E2F transcription factor 3	Protein binding, transcription factor, transcription initiation from Pol II promoter. Involved in cell cycle regulation, binds retinoblastoma protein (Rb). E2F family plays a crucial role in the control of cell cycle and action of tumor suppressor proteins and is also a target of the transforming proteins of small DNA tumor viruses.	0.049
AF217513	Hs.279905:73, Hs.283649:4	ANKT, nucleolar protein ANKT	clone HQ0310 PRO0310p1 nucleolar protein ANKT - no functional data	0.0504
AW513143	Hs.98367:8	ESTs	Expressed in uterus	0.0535
AJ245671	Hs.12844:73	EGFL6, EGF-like-domain; multiple 6	Cell cycle, oncogenesis, integrin ligand, extracellular space. Member of the epidermal growth factor (EGF) repeat superfamily; contains an EGF-like-domain. Expressed early during development, and its expression has been detected in lung and meningioma tumors.	0.0568
AA084248	Hs.85339:64	GPR39, G protein-coupled receptor 39	G-protein linked receptor, G-protein coupled receptor protein signaling pathway, integral plasma membrane protein.	0.19

Table 2 – Down regulated in OC

Accession number	UniGene Mapping	Gene symbol and title	Putative Function	P value
AI420582	Hs.136164:23	SE20-4, cutaneous T-cell lymphoma-associated tumor antigen se20-4se20-4	Cutaneous T-cell lymphoma-associated tumor antigen se20-4se20-4; differentially expressed nucleolar TGF-beta1 target protein (DENTT); also known as CDA1	0.0001
AI745249	Hs.23650:30	Homo sapiens, clone MGC:9889 IMAGE:3868330	Function unknown	0.0009
AI631024	Hs.24948:32, Hs.300445:4	SNCAIP, synuclein, alpha interacting protein (synphilin)	Cytoplasm, pathogenesis, protein binding. Synphilin-1; promotes formation of cytosolic inclusions in neurons (SNCAIP). Synuclein alpha interacting protein contains several protein-protein interaction domains and interacts with alpha synuclein in neurons. Mutations of SNCAIP have been linked to Parkinson disease.	0.001
M62397	Hs.1345:5	MCC, mutated in colorectal cancers	Receptor, signal transduction, tumor suppressor. Similar to the G protein-coupled m3 muscarinic acetylcholine receptor, MCC is a candidate for the putative colorectal tumor suppressor gene. The MCC gene product may be involved in early stages of colorectal neoplasia in both sporadic and familial tumors.	0.0161
AI694200	Hs.356620, Hs.227913:11	ESTs	Function unknown	0.0442

Table 3A – Serous OC

Accession number	UniGene Mapping	Gene symbol and title	Putative Function	P value
U62801	Hs.79361:65	KLK6, kallikrein 6 (neurosin, zyme)	Serine type peptidase, pathogenesis. Neurosin (protease M, zyme); a serine protease that cleaves amyloid precursor protein (APP). Growing evidence suggests that many kallikreins are implicated in carcinogenesis and some have potential as novel cancer and other disease biomarkers.	0.0159
D49441	Hs.155981:53	MSLN, mesothelin	Cell adhesion, cell surface antigen, membrane. Pre-pro-megakaryocyte potentiating factor. An antibody that reacts with ovarian cancers and mesotheliomas was used to isolate a cell surface antigen named mesothelin. Although the function of mesothelin is unknown, it may play a role in cellular adhesion and is present on mesothelium, mesotheliomas, and ovarian cancers.	0.147
X51630	Hs.1145:22,Hs.296851:1	WT1, Wilms tumor 1	Nucleus, transcription factor, transcription regulation. 4 Zn finger domains. Functions in kidney and gonad proliferation and differentiation. Mutations in this gene can be associated with the development of Wilms tumors in the kidney or with abnormalities of the genitourinary tract.	0.2938
AB018305	Hs.5378:149	SPON1, spondin 1, (f-spondin) extracellular matrix protein	Extracellular matrix protein. Very strongly similar to rat F-spondin (Rn.7546); may have a role in the growth and guidance of axons.	0.3394
AA433988	Hs.98502:8	MUC16, mucin 16, CA125	Mucin 16. Alias CA125 ovarian cancer antigen	0.6568

Table 3B – Serous OC ratios

Accession number	UniGene Mapping	Gene symbol and title	Putative Function	Ratio
M25809	Hs.64173	ATP6V1B1, ATPase, H+ transporting, lysosomal 56/58kD, V1 subunit B, isoform 1 (Renal tubular acidosis with deafness)	Subunit B1 (beta subunit) of a vacuolar-type H+-ATPase 1; apical proton pump that mediates distal nephron acid secretion	1062.30
AW959311	Hs.172012	DKFZP434J037: hypothetical protein DKFZp434J037	Function unknown	227.83
H16423	Hs.82685	Homo sapiens mRNA; cDNA DKFZp313F0317 (from clone DKFZp313F0317)	Function unknown	74.54
AI733848	Hs.71935	ZNF339, zinc finger protein 339	Zinc finger protein	55.13
AW055308	Hs.31803	NAC1, transcriptional repressor NAC1	Function unknown	52.63
AF034102	Hs.32951	SLC29A2, solute carrier family 29 (nucleoside transporters), member 2	Nitrobenzylthioinosine-insensitive equilibrative nucleoside transporter 2; may act in the uptake of purine and pyrimidine nucleosides	44.34
AI791905	Hs.95549	FLJ20273: RNA-binding protein	Contains four RNA recognition motifs (RRM, RBD, or RNP)	43.21
AW296454	Hs.24743	FLJ20171: hypothetical protein FLJ20171	Contains three RNA recognition motifs (RRM, RBD, or RNP)	38.91
Z43989	Hs.82141	Human clone 23612 mRNA sequence	Function unknown	37.89
AL043980	Hs.7886	PELI1, pellino homolog 1 (Drosophila)	Pellino protein	35.20
BE514982	Hs.38991	S100A2, S100 calcium binding protein A2	S100 calcium-binding protein A2; interacts with target proteins to link extracellular stimuli and cellular responses; member of the S100 tissue/cell specific Ca2+-binding protein family	34.53
		Target Exon	Function unknown	34.02
AI811807	Hs.108646	Homo sapiens cDNA FLJ12534 fis, clone NT2RM4000244	Function unknown	32.34
U90441	Hs.3622	P4HA2, procollagen-proline, 2-oxoglutarate 4-dioxygenase (proline 4-hydroxylase), alpha polypeptide II	Alpha 2 subunit of prolyl 4-hydroxylase; catalyzes the formation of 4-hydroxyproline in collagens	32.24
T98226	Hs.171952	OCLN, occludin	This gene encodes an integral membrane protein which is located at tight junctions. This protein may be involved in the formation and maintenance of the tight junction.	31.56
R35343	Hs.24968		Human DNA sequence from clone RP1-233G16 on chromosome Xq22.1-23. Contains the 5' part of a novel gene, ESTs, STSs, GSSs and a putative CpG island	31.22
BE247295	Hs.78452	SLC20A1, solute carrier family 20 (phosphate transporter), member 1	Sodium-dependent phosphate symporter; acts as a cell-surface receptor for gibbon ape leukemia virus	30.16

AB037734	Hs.4993	PCDH19, protocadherin	Protocadherin	29.90
		C5000394':gi|12737280|ref|XP_006682.2| keratin 18 Homo sapiens]||6633	Function unknown	29.30
AF212223	Hs.25010	Homo sapiens BM025 mRNA, complete cds	Function unknown	28.85
AA902656	Hs.21943	NIF3L1, NIF3 (Ngg1 interacting factor 3, S.pombe homolog)-like 1	Amyotrophic lateral sclerosis 2 (juvenile) chromosome region, candidate 1	27.73
X14008	Hs.234734	Human lysozyme gene (EC 3.2.1.17)	Lysozyme	27.66
AA570256		LOC116238: hypothetical protein BC014072	Function unknown	27.52
AA137152	Hs.286049	PSA, phosphoserine aminotransferase	The protein encoded by this gene is likely a phosphoserine aminotransferase, based on similarity to proteins in mouse, rabbit, and Drosophila. Alternative splicing of this gene results in two transcript variants encoding different isoforms.	25.57
BE621807		TM4SF1, transmembrane 4 superfamily member 1	L6 antigen; member of the transmembrane 4 superfamily (TM4SF)	25.40
AB041036	Hs.57771	KLK11, kallikrein 11	Trypsin-like serine protease; has serine protease activity	25.05
F13386	Hs.7888	Homo sapiens clone 23736 mRNA sequence	Function unknown	22.50
AA158177	Hs.118722	FUT8, fucosyltransferase 8 (alpha (1,6) fucosyltransferase)	N-linked glycosylation, oligosaccharide biosynthesis, glycoprotein 6-alpha-L-fucosyltransferase. Alpha(1,6)fucosyltransferase (GDP-L-Fuc:N-acetyl-beta-D-glucosaminide:alpha1-6 fucosyltransferase); transfers fucose to N-linked type complex glycopeptides from GDP-Fuc; functions in asparagine-linked glycoprotein oligosaccharide synthesis	21.90
BE267045	Hs.75064	TBCC, tubulin-specific chaperone c	Tubulin-specific chaperone c; cofactor in the folding pathway of beta-tubulin, mediates the release of beta-tubulin polypeptides committed to the native state	21.49
		NM_005936:Homo sapiens myeloid/lymphoid or mixed-lineage leukemia (trithorax (Drosophila) homolog); translocated to, 4 (MLLT4), mRNA.	Function unknown	20.46
AA150864	Hs.790	MGST1, microsomal glutathione S-transferase 1	Microsome, glutathione transferase. Microsomal glutathione S-transferase; catalyzes the conjugation of glutathione to electrophilic compounds; member of a family of detoxication enzymes.	20.35
AW955632	Hs.66666	EST367702 MAGE resequences, MAGD Homo sapiens cDNA, mRNA sequence	Function unknown	20.26
AW837046	Hs.6527	QV1-LT0037-150200-069-e09 LT0037 Homo sapiens cDNA, mRNA sequence	Function unknown	19.60
AA286887	Hs.24724	MFHAS1, malignant fibrous histiocytoma amplified sequence 1	The primary structure of its product includes an ATP/GTP-binding site, three leucine zipper domains, and a leucine-rich tandem repeat, which are structural or functional elements for interactions among proteins related to the cell cycle, and which suggest that overexpression might be oncogenic with respect to MFH.	19.16
AW401864	Hs.18720	PDCD8: programmed cell death 8 (apoptosis-inducing factor)	Mitochondrial apoptosis-inducing factor; flavoprotein inducing chromatin condensation and DNA fragmentation	19.01

AA196241	Hs.73980	zp98f03.r1 Stratagene muscle 937209 Homo sapiens cDNA clone IMAGE:628253 5' similar to gb:M19309 TROPONIN T, SLOW SKELETAL MUSCLE ISOFORMS (HUMAN);, mRNA sequence	Function unknown	18.82
NM_004998	Hs.82251	MYO1E, myosin IE	Highly similar to class I myosin; may bind proline-rich peptides; contains an Src homology 3 (SH3) and myosin head domain (motor domain)	18.62
AW873704	Hs.320831	C20orf72: chromosome 20 open reading frame 72	Function unknown	18.19
AW361666	Hs.49500	KIAA0746: KIAA0746 protein	Function unknown	18.05
BE174595	Hs.366	PTS, 6-pyruvoyltetrahydropterin synthase	6-Pyruvoyltetrahydropterin synthase; synthesizes tetrahydrobiopterin, activity requires sepiapterin reductase, Mg2+, and NADPH	17.28
M31669	Hs.1735	Human inhibin beta-B-subunit gene, exon 2, and complete cds	Function unknown	16.24
AK001714	Hs.95744	FLJ10852, hypothetical protein similar to ankyrin repeat-containing priotein AKR1	May be involved in protein-protein interactions; has five ankyrin repeats and a DHHC-type zinc finger or NEW1 domain	16.09
AU076517	Hs.184276	AU076517 Sugano cDNA library Homo sapiens cDNA clone ColF3365 similar to 5'-end region of Homo sapiens ezrin-radixin-moesin binding phosphoprotein-50 mRNA, mRNA sequence	Function unknown	16.05
NM_006456	Hs.288215	STHM, sialyltransferase	Low similarity to beta-galactosidase a-2,3-sialytransferase SIAT4B; member of the sialyltransferase family	15.93
BE148235	Hs.193063	Homo sapiens cDNA FLJ14201 fis, clone NT2RP3002955	Function unknown	15.91
AV653729	Hs.8185	SQRDL: sulfide dehydrogenase like (yeast)	Sulfide dehydrogenase like	15.35
AL119671	Hs.1420	FGFR3, fibroblast growth factor receptor 3 (achondroplasia, thanatophoric dwarfism)	Fibroblast growth factor receptor 3; receptor tyrosine kinase that binds acidic and basic FGF	14.62
AA393071	Hs.182579	LAP3, leucine aminopeptidase	Leucine aminopeptidase	14.60
AL048753	Hs.303649	CCL2, chemokine (C-C motif) ligand 2	Cytokine A 2; chemotactic factor for monocytes	14.37
AI868872	Hs.282804	CP, ceruloplasmin (ferroxidase)	Ceruloplasmin; ferrous oxidase, binds copper in plasma and maintains iron homeostasis	14.07
NM_004419	Hs.2128	DUSP5, dual specificity phosphatase 5	Mitogen inducible dual specificity protein phosphatase 5; dephosphorylates extracellular signal-regulated kinase	14.05
AW969587	Hs.86366	EST381664 MAGE resequences, MAGK Homo sapiens cDNA, mRNA sequence	Function unknown	13.75
AW161449	Hs.72290	WNT7A, wingless-type MMTV integration site family, member 7A	Very strongly similar to murine Wnt7a; may have a role in limb development and sexual dimorphism; member of the Wnt family of cell signalling proteins	13.48
BE409838	Hs.194657	CDH1, cadherin 1, type 1, E-cadherin (epithelial)	E-cadherin (uvomorulin); Ca2+-dependent glycoprotein, mediates cell-cell interactions in	12.92

			epithelial cells	
BE540274	Hs.239	FOXM1, forkhead box M1	Cell-cycle regulated HNF-3/fork head; a transcriptional regulator	12.86
AF022375	Hs.73793	VEGF, vascular endothelial growth factor	Vascular endothelial growth factor; induces endothelial cell proliferation and vascular permeability	12.79
AW369278	Hs.23412	FLJ20160: hypothetical protein FLJ20160	Function unknown	12.73
AF147204	Hs.89414	CXCR4, chemokine (C-X-C motif), receptor 4 (fusin)	CXC chemokine receptor (fusin); G protein-coupled receptor binds CXC cytokines, mediates intracellular calcium flux	12.56
BE242818	Hs.311609	DDX39, DEAD/H (Asp-Glu-Ala-Asp/His) box polypeptide 39	Strongly similar to human D6S81E; member of the DEAD/H box ATP-dependent RNA helicase family	12.43
NM_014791	Hs.184339	MELK, maternal embryonic leucine zipper kinase	Leucine zipper kinase	12.25
U38847	Hs.151518	TARBP1, TAR (HIV) RNA binding protein 1	Binds to the HIV-1 TAR RNA regulatory element, may function alone or with HIV-1 Tat to disengage RNA polymerase II during transcriptional elongation; has a leucine zipper	12.22
AW953575	Hs.303125	EST365645 MAGE resequences, MAGC Homo sapiens cDNA, mRNA sequence	Function unknown	12.21
AI949095	Hs.67776	ESTs, Weakly similar to T22341 hypothetical protein F47B8.5 - Caenorhabditis elegans [C.elegans]	Homo sapiens, clone IMAGE:5455669, mRNA, partial cds	12.08
BE274530	Hs.273333	FLJ10986, hypothetical protein FLJ10986	Member of the FGGY carbohydrate kinase family	11.75
AB020676	Hs.21543	KIAA0869 protein	Function unknown	11.73
		Target Exon	Function unknown	11.69
H48299	Hs.26126:33	CLDN10, claudin 10	Cell adhesion, integral plasma membrane protein, tight junction.	11.67
T34530	Hs.4210	Homo sapiens cDNA FLJ13069 fis, clone NT2RP3001752	Function unknown	11.50
AW088642	Hs.97984	SOX17, SRY (sex determining region Y)-box 17	SRY-related HMG-box transcription factor SOX17	11.42
AA737033	Hs.7155	Homo sapiens, clone IMAGE:4428577, mRNA, partial cds	Function unknown	10.79
AA433988	Hs.98502:8	MUC16, mucin 16, CA125	Mucin 16. Alias CA125 ovarian cancer antigen	10.52
H91282	Hs.286232	Homo sapiens cDNA: FLJ23190 fis, clone LNG12190	Function unknown	10.50
AW005054	Hs.47883	LOC57118: CamKI-like protein kinase	CamKI-like protein kinase; granulocyte-specific protein kinase that activates ERK/MAP kinase activity; similar to Ca(2+)-calmodulin-dependent kinase I (CamKI)	10.49
X69699	Hs.73149	PAX8, paired box gene 8	Member of the paired domain family of nuclear transcription factors; may be involved in the ribosome assembly, required for normal thyroid development	10.39
AW382987	Hs.88474:42	Homo sapiens cDNA, mRNA sequence	Function unknown	10.21

AW957446	Hs.301711	Homo sapiens, clone MGC:23936 IMAGE:3838595, mRNA, complete cds	Function unknown	10.12
AA361562	Hs.178761	POH1: 26S proteasome-associated pad1 homolog	Ubiquitin-dependent protein degradation	10.01
AA834626		RAD54L, RAD54 (S.cerevisiae)-like	Has likely roles in mitotic and meiotic DNA recombination and repair; member of SNF2/SWI2 family of DNA-dependent ATPases	9.85
AI878927	Hs.79284	MEST, mesoderm specific transcript (mouse) homolog	Mesoderm specific protein; member of the alpha/beta hydrolase fold family	9.83
AW074266	Hs.23071	LOC85439: stonin 2	Stonin 2	9.74
NM_000947	Hs.74519	PRIM2A, primase, polypeptide 2A (58kD)	Subunit of DNA primase polypeptide 2A; part of the DNA polymerase alpha-primase complex	9.72
NM_006187	Hs.56009	OAS3, 2'-5'-oligoadenylate synthetase 3 (100 kD)	Member of the 2'5'-oligoadenylate synthetase family	9.68
AW276858	Hs.81256	S100A4, S100 calcium binding protein A4 (calcium protein, calvasculin, metastasin, murine placental homolog)	Calcyclin (metastasis-associated) (S100 calcium-binding protein A4); interacts with targets to link extracellular stimuli and cellular responses; member of the S100 family of tissue-specific calcium-binding proteins	9.66
T18997	Hs.180372	LOC139231: hypothetical protein BC016683	Function unknown	9.49
AA262294	Hs.180383	DUSP6, dual specificity phosphatase 6	Dual specificity protein phosphatase 6; selectively dephosphorylates and inactivates MAP kinase	9.48
AA220238	Hs.94986	RPP38: ribonuclease P (38kD)	Nucleus, ribonuclease P. Subunit p38 of ribonuclease P ribonucleoprotein; processes 5' ends of precursor tRNAs	9.41
AW505308	Hs.75812	PCK2, phosphoenolpyruvate carboxykinase 2 (mitochondrial)	Phosphoenolpyruvate carboxykinase 2; forms phosphoenolpyruvate by decarboxylation of oxaloacetate at the rate-limiting step of gluconeogenesis	9.38
AI186431	Hs.296638	PLAB: prostate differentiation factor	Macrophage inhibitory cytokine; member of a subgroup of the TGF-beta superfamily	9.12
AI095718	Hs.135015	Homo sapiens cDNA FLJ40906 fis, clone UTERU2004698, highly similar to Mus musculus mRNA for thrombospondin type 1 domain	Function unknown	9.04
W70171	Hs.75939	UMPK, uridine monophosphate kinase	The protein encoded by this gene catalyzes the phosphorylation of uridine monophosphate to uridine diphosphate. This is the first step in the production of the pyrimidine nucleoside triphosphates required for RNA and DNA synthesis. In addition, an allele of this gene may play a role in mediating nonhumoral immunity to Hemophilus influenzae type B.	8.97
AI580935	Hs.105698	Homo sapiens cDNA FLJ31553 fis, clone NT2RI2001178	Function unknown	8.90
AB040914	Hs.278628	ShrmL: Shroom-related protein	Shroom-related protein	8.87
AU076611	Hs.154672	MTHFD2, methylene tetrahydrofolate dehydrogenase (NAD+ dependent), methenyltetrahydrofolate cyclohydrolase	NAD-dependent methylene tetrahydrofolate dehydrogenase-cyclohydrolase; may provide formyltetrahydrofolate for formylmethionyl tRNA synthesis; involved in initiation of mitochondrial protein synthesis	8.71

AI089660	Hs.323401	LOC84661: dpy-30-like protein	dpy-30-like protein	8.71
D13666	Hs.136348:228,Hs.80988:2	OSF-2: osteoblast specific factor 2 (fasciclin I-like)	Cell adhesion, skeletal development. Putative bone adhesion protein; similar to the insect protein fasciclin I	8.64
AI798863	Hs.87191	ESTs	Function unknown	8.52
U78093	Hs.15154	SRPX, sushi-repeat-containing protein, X chromosome	Putative membrane protein with short consensus repeat (sushi) domains	8.51
AI669760	Hs.188881	ESTs	Function unknown	8.37
AI375726	Hs.279918	MGC2198: hypothetical protein MGC2198	Function unknown	8.37
AW271106	Hs.133294	ESTs	Function unknown	8.30
AK001782	Hs.15093	HSPC195: hypothetical protein HSPC195	Function unknown	8.18
AF019226	Hs.8036	RAB3D,RAB3D, member RAS oncogene family	GTP-binding protein; may be involved in vesicle transport; member of the RAB family of small GTPases	7.94
AW968343	Hs.24255	LOC150696: prominin-related protein	Prominin-related protein	7.90
AF111856	Hs.105039	SLC34A2, solute carrier family 34 (sodium phosphate), member 2	Sodium-dependent phosphate transporter; member of the renal type II co-transporter family	7.87
AA863360	Hs.26040	Homo sapiens, clone MGC:40051 IMAGE:5243005, mRNA,complete cds	Function unknown	7.75
NM_005764	Hs.271473	DD96: epithelial protein up-regulated in carcinoma, membrane associated protein 17	Up-regulated in malignant epithelial cells of renal cell carcinomas, and in carcinomas of colon, breast and lung	7.75
AW360901	Hs.183047	MGC4399: mitochondrial carrier protein	Mitochondrial carrier protein MGC4399	7.71
AL353944	Hs.50115	Homo sapiens mRNA; cDNA DKFZp761J1112 (from clone DKFZp761J1112)	Function unknown	7.69
H59799	Hs.42644	TXNL2, thioredoxin-like 2	Member of the thioredoxin family; has region of moderate similarity to glutaredoxin-like proteins	7.65
NM_002984	Hs.75703	CCL4, chemokine (C-C motif) ligand 4	Cytokine A4	7.64
AA642452	Hs.130881	BCL11A, B-cell CLL/lymphoma 11A (zinc finger protein)	May bind nucleic acids; contains three C2H2 type zinc finger domains	7.61
AA789081	Hs.4029	GAS41: glioma-amplified sequence-41	Similar to the transcription factors AF-9 and ENL	7.46
H13032	Hs.103378	MGC11034, hypothetical protein MGC11034	Function unknown	7.42
BE384836	Hs.3454	KIAA1821: KIAA1821 protein	KIAA1821 protein	7.40
AW067800	Hs.155223	STC2, stanniocalcin 2	Stanniocalcin 2; may regulate metal ion homeostasis and inhibits phosphate uptake	7.36
T55979	Hs.115474	RFC3, replication factor C (activator 1) 3 (38kD)	Subunit of replication factor C (activator 1) 3; activator of DNA polymerases	7.35
AJ278016	Hs.55565	ANKRD3, ankyrin repeat domain 3	Ortholog of mouse protien kinase C-associated kinase, putative gene, ankirin like,	7.25

			possible dual-specificity Ser/Thr/Tyr kinase domain	
		NM_025080:Homo sapiens hypothetical protein FLJ22316 (FLJ22316), mRNA. VERSION NM_025079.1 GI:13376631	Function unknown	7.22
AA084248	Hs.85339:64	GPR39, G protein-coupled receptor 39	GPR39, G protein-coupled receptor 39	7.15
BE620738	Hs.173125	PPIF, peptidylprolyl isomerase F (cyclophilin F)	Cyclophilin F (peptidylprolyl isomerase F); binds the immunosuppressant drug cyclosporin A	7.06
AF072873	Hs.114218	FZD6, frizzled (Drosophila) homolog 6	Frizzled-6; may function in tissue polarity, development and carcinogenesis; similar to frizzled receptor family, has seven transmembrane domains	7.04
AA852773	Hs.334838	KIAA1866 protein	KIAA1866 protein	6.99
R07566	Hs.73817	CCL3, chemokine (C-C motif) ligand 3	Macrophage inflammatory protein 1 alpha; chemokine	6.98
NM_005211	Hs.174142	CSF1R, colony stimulating factor 1 receptor, formerly McDonough feline sarcoma viral (v-fms) oncogene homolog	Macrophage colony stimulating factor tyrosine kinase receptor; involved in regulation of growth and differentiation of myeloid cells	6.79
AI752666	Hs.76669	NNMT, nicotinamide N-methyltransferase	Nicotinamide N-methyltransferase; catalyzes the N-methylation of nicotinamide and other pyridines, structurally-related drugs and xenobiotics	6.52
AF182294	Hs.241578	LOC51691: U6 snRNA-associated Sm-like protein LSm8	Member of the Sm family; core constituent of snRNP complexes	6.50
AA457211	Hs.8858	BAZ1A, bromodomain adjacent to zinc finger domain, 1A	May bind DNA and act as a chromatin-mediated transcriptional regulator; contains a bromodomain and a PHD-finger	6.48
W40262	Hs.146310	zc79f02.s1 Pancreatic Islet Homo sapiens cDNA clone IMAGE:328539 3', mRNA sequence	Function unknown	6.47
AB033091	Hs.74313	KIAA1265 protein	Function unknown	6.45
AA292998	Hs.163900	ESTs, Highly similar to winged helix/forkhead transcription factor [Homo sapiens] [H.sapiens]	Function unknown	6.36
BE613269	Hs.21893	DKFZp761N0624: hypothetical protein DKFZp761N0624	Function unknown	6.35
H25836	Hs.301527	ESTs, Moderately similar to unknown [Homo sapiens] [H.sapiens]	Function unknown	6.27
AL037228	Hs.82043	NUDT5, nudix (nucleoside diphosphate linked moiety X)-type motif 5	NDP-sugar hydrolase; converts ADP-ribose to AMP or ribose 5-phosphate; contains a MutT motif	6.25
AV662037	Hs.124740	FLJ30532: hypothetical protein FLJ30532	Function unknown	6.21
AI674383	Hs.22891	wc38h08.x1 NCI_CGAP_Pr28 Homo sapiens cDNA clone IMAGE:2320959 3', mRNA sequence	Function unknown	6.20

AW342140	Hs.182545	ESTs, Weakly similar to POL2_MOUSE Retrovirus-related POL polyprotein [Contains: Reverse transcriptase ; Endonuclease] [M.musculus]	Function unknown	6.18
BE560135	Hs.5232	HSPC125, HSPC125 protein	Function unknown	6.17
BE409857	Hs.69499	HSPC132: hypothetical protein HSPC132	Moderately similar to a region of S. cerevisiae Ykl053c-ap	6.16
AW972542	Hs.289008	LOC116150: hypothetical protein, MGC:7199	Function unknown	6.16
AI523755	Hs.59236	DKFZP434L0718: hypothetical protein DKFZp434L0718	Function unknown	6.16
NM_014056	Hs.7917	DKFZP564K247: DKFZP564K247 protein	Function unknown	6.08
AI857607	Hs.181301	CTSS, cathepsin S	Cathepsin S; lysosomal cysteine (thiol) protease that cleaves elastin	6.04
AW247529	Hs.6793	PAFAH1B3, platelet-activating factor acetylhydrolase, isoform Ib, gamma subunit (29kD)	Platelet-activating factor acetylhydrolase gamma; may play a role in brain development	5.98
AK000868	Hs.5570	Homo sapiens cDNA FLJ10006 fis, clone HEMBA1000168, weakly similar to CYLICIN I	Function unknown	5.92
AF053551	Hs.31584	MTX2, metaxin 2	Very strongly similar to murine metaxin 2 (Mm.12941); may be involved in mitochondrial protein import	5.91
AI538613	Hs.298241	TMPRSS3, Transmembrane protease, serine 3	The encoded protein contains a serine protease domain, a transmembrane domain, a LDL receptor-like domain, and a scavenger receptor cysteine-rich domain. This gene was identified as a tumor associated gene that is overexpressed in ovarian tumors.	5.86
U48508	Hs.89631	Human skeletal muscle ryanodine receptor gene (RYR1), exons 103, 104, 105, 106, and complete cds	Function unknown	5.86
T69387	Hs.76364	AIF1, allograft inflammatory factor 1	Allograft inflammatory factor 1; cytokine inducible protein associated with vascular injury	5.86
AC005954	Hs.25527	Homo sapiens chromosome 19, cosmid R28784, complete sequence	Function unknown	5.86
AB037805	Hs.88442	KIAA1384 protein	Function unknown	5.84
AL031427	Hs.40094	Human DNA sequence from clone 167A19 on chromosome 1p32.1-33. Contains three genes for novel proteins, the DIO1 gene for type I iodothyronine deiodinase (EC 3.8.1.4, TXDI1, ITDI1) and an HNRNP A3 (Heterogenous Nuclear Ribonucleoprotein A3, FBRNP) pseudogene.	Function unknown	5.83
AA340864	Hs.278562	CLDN7, claudin 7	Similar to murine Cldn7; may be an integral membrane protein	5.76
X89984	Hs.211563	BCL7A, B-cell CLL/lymphoma 7A	Similar to the actin-binding protein caldesmon; serine-rich	5.74
AI355761	Hs.242463	qt94a11.x1 NCI_CGAP_Co14 Homo sapiens cDNA clone IMAGE:1962908 3' similar to gb:X74929 KERATIN, TYPE II CYTOSKELETAL 8 (HUMAN);, mRNA sequence	Function unknown	5.73

AA376409	Hs.10862	Homo sapiens cDNA: FLJ23313 fis, clone HEP11919	Function unknown	5.71
AA310162	Hs.169248	HCS: cytochrome c	Somatic cytochrome c (heart cytochrome c)	5.67
AW015534	Hs.217493	ANXA2, annexin A2	Annexin II (lipocortin-2); enhances osteoclast formation and bone resorption; member of the annexin protein family	5.64
AA326108	Hs.53631:82	BHLHB3: basic helix-loop-helix domain containing, class B, 3	Basic helix-loop-helix (bHLH) transcription factors (e.g., DEC1, also called BHLHB2; 604256) are related to Drosophila hairy/enhancer of split proteins. They are involved in the control of proliferation and development during differentiation, particularly in neurons.	5.64
AA120865	Hs.23136	ESTs, Highly similar to THYA_HUMAN Prothymosin alpha [H.sapiens]	Function unknown	5.62
AK000517	Hs.6844	NALP2: NALP2 protein	Protein with low similarity to murine Op1	5.54
Z36842	Hs.57548	H.sapiens (xs85) mRNA, 209bp	Function unknown	5.53
AA831552	Hs.268016	Homo sapiens cDNA: FLJ21243 fis, clone COL01164		5.50
AL137578	Hs.27607	Homo sapiens mRNA; cDNA DKFZp564N2464 (from clone DKFZp564N2464)	Function unknown	5.50
AA316181	Hs.61635	STEAP, six transmembrane epithelial antigen of the prostate	Six transmembrane epithelial antigen of the prostate; prostate-specific cell-surface antigen	5.46
X03635	Hs.1657	ESR1, estrogen receptor 1	Estrogen receptor; nuclear receptor transcription factor activated by ligand-binding, involved in hormone-mediated inhibition of gene expression	5.42
AI557280	Hs.184270	PT2.1_15_G11.r tumor2 Homo sapiens cDNA 3', mRNA sequence	Function unknown	5.41
AW248508	Hs.279727	Homo sapiens cDNA FLJ14035 fis, clone HEMBA1004638	Function unknown	5.40
N90866	Hs.276770	CDW52, CDW52 antigen (CAMPATH-1 antigen)	CAMPATH-1 antigen; GPI-anchored protein	5.39
U83115	Hs.161002	AIM1, absent in melanoma 1	Member of the beta gamma-crystallin superfamily of proteins; interactions with the cytoskeleton	5.35
AB007860	Hs.12802	DDEF2, development and differentiation enhancing factor 2	GTPase-activating protein; interacts with members of the Arf and Src family	5.35
Z46223	Hs.176663	H.sapiens DNA for immunoglobulin G Fc receptor IIIB	Immunoglobulin G Fc receptor	5.31
BE264974	Hs.6566	TRIP13, thyroid hormone receptor interactor 13	Interacts with ligand binding domain of thyroid hormone receptor and with human papillomavirus type 16 (HPV16) E1	5.30
AA194422	Hs.22564	MYO6, myosin VI	Motor, hearing, myosin ATPase, structural protein. Class 6 myosin; motor protein; very strongly similar to murine Myo6	5.27

AF134157	Hs.169487	MAFB, v-maf musculoaponeurotic fibrosarcoma oncogene homolog B (avian)	Very strongly similar to murine Krml; may function as a basic domain-leucine zipper transcription factor	5.25
AA232119	Hs.16085	SH120: putative G-protein coupled receptor	putative G-protein coupled receptor	5.25
W58353	Hs.285123	OSBPL10, oxysterol binding protein-like 10	Member of the oxysterol-binding protein (OSBP) family; may bind oxygenated derivatives of cholesterol	5.21
AW167128	Hs.231934	ESTs, Weakly similar to A57717 transcription factor EC2 - human [H.sapiens]	Function unknown	5.19
U70370	Hs.84136	PITX1, paired-like homeodomain transcription factor 1	Member of the homeodomain family of DNA binding proteins; may regulate gene expression and control cell differentiation	5.18
N55669	Hs.333823	MRPL13, mitochondrial ribosomal protein L13	Protein of the large 60S ribosomal subunit	5.17
BE298446	Hs.305890	BCL2L1, BCL2-like 1	BCL2-related protein; alternative form bcl-xlong inhibits apoptosis and bcl-xshort induces apoptosis	5.17
AW136551	Hs.181245	Homo sapiens cDNA FLJ12532 fis, clone NT2RM4000200	Function unknown	5.15
AW250380	Hs.109059	HGS, hepatocyte growth factor-regulated tyrosine kinase substrate	Zinc-finger protein; interacts with STAM, undergoes tyrosine phosphorylation in response to IL2, CSF2, or HGF	5.13
AW002565	Hs.124660	Homo sapiens cDNA: FLJ21763 fis, clone COLF6967	Function unknown	5.13
AI697274	Hs.105435	GMDS, GDP-mannose 4,6-dehydratase	GDP-mannose-4,6-dehydratase; epimerase converts GDP-mannose to GDP-mannose-4-keto-6-D-deoxymannose, plays a role in the synthesis of fucosylated oligosaccharides	5.11
NM_003878	Hs.78619	GGH, gamma-glutamyl hydrolase (conjugase, folylpolygammaglutamyl hydrolase)	Gamma-glutamyl hydrolase; has greater exopeptidase activity on methotrexate pentaglutamate than on diglutamate	5.11
AF052112	Hs.12540	LYPLA1, lysophospholipase I	Lysophospholipid-specific lysophospholipase 1; hydrolyzes lysophosphatidyl choline	5.09
AV654694	Hs.82316	IFI44, interferon-induced protein 44	Member of the family of interferon-alpha/beta inducible proteins; may mediate the antiviral action of interferon	5.09
R24601		Homo sapiens adenylosuccinate synthetase isozyme (ADSS) mRNA, complete cds	Adenylosuccinate synthetase	5.07
BE019020	Hs.85838	Homo sapiens cDNA clone IMAGE:2963945 5' similar to TR:O15427 O15427 MONOCARBOXYLATE TRANSPORTER. ;, mRNA sequence	Function unknown	5.04
AW163799	Hs.198365	BPGM, 2,3-bisphosphoglycerate mutase	2,3-bisphosphoglycerate mutase; has synthase, mutase, and phosphatase activities, controls 2,3-diphosphoglycerate metabolism, which is an effector for haemoglobin	5.04
AA278921	Hs.1908	PRG1, proteoglycan 1, secretory granule	Secretory granule proteoglycan 1	5.02
NM_003728	Hs.19126	SCAP1, src family associated phosphoprotein 1	Src kinase-associated phosphoprotein; acts as an adaptor protein; contains a pleckstrin homology domain and an SH3 domain	5.02

AA281167	Hs.111911	ESTs, Weakly similar to T06291 extensin homolog T9E8.60 - Arabidopsis thaliana [A.thaliana]	Function unknown	5.02
		C9000306*:gi|12737280|ref|XP_006682.2| keratin 18 [Homo sapiens]||6633	Function unknown	5.01
AF098158	Hs.9329	C20orf1, chromosome 20 open reading frame 1	Proliferation-associated nuclear protein; associates with the spindle pole and mitotic spindle during mitosis	5.00
AA101043	Hs.151254:19	KLK7, kallikrein 7 (chymotryptic; stratum corneum)	Epidermal differentiation. Stratum corneum chymotryptic enzyme; serine protease. Growing evidence suggests that many kallikreins are implicated in carcinogenesis and some have potential as novel cancer and other disease biomarkers. Thought to be involved in the proteolysis of intercellular cohesive structures preceding desquamation, which is the shedding of the outermost layer of the epidermis.	4.87
AF017986	Hs.31386:185	Homo sapiens secreted apoptosis related protein 1 (SARP1) mRNA, partial cds.	Function unknown	4.12
AW960564	Hs.3337:137	TM4SF1: transmembrane 4 superfamily member 1	Pathogenesis, plasma membrane, cell proliferation, N-linked glycosylation, integral membrane protein, integral plasma membrane protein. L6 antigen; member of the transmembrane 4 superfamily (TM4SF). The proteins mediate signal transduction events that play a role in the regulation of cell development, activation, growth and motility. This encoded protein is a cell surface antigen and is highly expressed in different carcinomas.	3.62
W29092	Hs.7678:40	CRABP1 Cellular retinoic acid binding protein 1	Cytoplasm, retinoid binding, signal transduction, developmental processes. Cellular retinoic acid-binding protein 1; may be involved in delivering retinoic acid to the nucleus, assumed to play an important role in retinoic acid-mediated differentiation and proliferation processes.	3.34
H93366	Hs.7567:84	Homo sapiens cDNA: FLJ21962 fis, clone HEP05564	Function unknown	3.29
D49441	Hs.155981:53	MSLN, mesothelin	Cell adhesion, cell surface antigen, membrane. Pre-pro-megakaryocyte potentiating factor. An antibody that reacts with ovarian cancers and mesotheliomas was used to isolate a cell surface antigen named mesothelin. Although the function of mesothelin is unknown, it may play a role in cellular adhesion and is present on mesothelium, mesotheliomas, and ovarian cancers.	3.14
AA214228	Hs.127751:21, Hs.78006:5	C20orf180: chromosome 20 open reading frame 180	Region of high similarity to tyrosine-phosphorylated protein DOK1	2.99
M31126	Hs.272620:1	PSG9: pregnancy specific beta-1-glycoprotein 9	Pregnancy, extracellular, plasma glycoprotein. Member of the pregnancy-specific glycoprotein (PSG) and CEA families.	2.82
U62801	Hs.79361:65	KLK6, kallikrein 6 (neurosin, zyme)	Serine type peptidase, pathogenesis. Neurosin (protease M, zyme); a serine protease that cleaves amyloid precursor protein (APP). Growing evidence suggests that many kallikreins are implicated in carcinogenesis and some have potential as novel cancer and other disease biomarkers.	2.77
AK001536	Hs.285803:6	Homo sapiens cDNA FLJ12852 fis, clone NT2RP2003445	Function unknown	2.73
NM_014767	Hs.74583:140	KIAA0275: KIAA0275 gene product	Function unknown	2.72

NM_000699	Hs.75733:129, Hs.278399:100, Hs.274376:1	AMY2A: amylase, alpha 2A; pancreatic	Alpha-amylase, extracellular space, carbohydrate metabolism. Pancreatic alpha-amylase 2A (1,4-alpha-D-glucan glucanohydrolase); cleaves internal a-1,4 bonds between glucose monomers to digest starch.	2.71
AA430348	Hs.288837:40	Homo sapiens cDNA FLJ12927 fis, clone NT2RP2004743	Function unknown	2.69
X51630	Hs.1145:22, Hs.296851:1	WT1, Wilms tumor 1	Nucleus, transcription factor, transcription regulation. 4 Zn finger domains. Functions in kidney and gonad proliferation and differentiation. Mutations in this gene can be associated with the development of Wilms tumors in the kidney or with abnormalities of the genitourinary tract.	2.58
BE393948	Hs.50915:17	KLK5, kallikrein 5	Serine type peptidase, epidermal differentiation, extracellular space. Stratum corneum tryptic enzyme (kallikrein-like protein); may function in epidermal stratum corneum desquamation and turnover. Expression in prostate cancer negatively correlated with cancer aggressiveness (Yousef 2002)	2.34
NM_002776	Hs.69423:46	KLK10, kallikrein 10	Putative serine protease. Expressed in normal breast tissue and benign lesions, with loss of expression during tumor progression (Dhar 2001). SNPs associated with prostate, breast, testicular, and ovarian cancers (Bharaj 2002).	2.24
NM_000954	Hs.8272:294	PTGDS: prostaglandin D2 synthase (21kD, brain)	Membrane, prostaglandin-D synthase. Glutathione-independent prostaglandin D2 synthase; membrane associated, catalyzes synthesis of prostaglandin D; member of the lipocalin family of transporters.	2.15
AB029000	Hs.70823:109, Hs.297970:48	KIAA1077: sulfatase FP	Function unknown	2.04
AL044315	Hs.173094:70	KIAA1750; KIAA1750 protein	Function unknown	0.95
AA334592	Hs.79914:337	LUM: lumican	Vision, proteoglycan, extracellular matrix, cartilage condensation, extracellular matrix glycoprotein. Member of the specialized collagens and SLRP_family	0.93
S79895	Hs.83942:248	CTSK: cathepsin K (pycnodysostosis)	Lysosome, cathepsin K, cysteine-type peptidase, proteolysis and peptidolysis. Cathepsin K (cathepsin O), a cysteine (thiol) protease; involved in bone remodeling and reabsorption	0.91
AI091195	Hs.65029:120	Homo sapiens cDNA clone IMAGE:1566910 3', mRNA sequence	Function unknown	0.91
AF026692	Hs.105700:83, Hs.278611:3	SFRP4: secreted frizzled-related protein 4	Member of the SFRP family that contains a cysteine-rich domain homologous to the putative Wnt-binding site of Frizzled proteins. SFRPs act as soluble modulators of Wnt signaling. The expression of SFRP4 in ventricular myocardium correlates with apoptosis related gene expression.	0.73
AI683243	Hs.97258:15	ESTs, Moderately similar to S29539 ribosomal protein L13a, cytosolic	Function unknown	-2.96
AI267700	Hs.111128:7	Homo sapiens, clone IMAGE:4106329, mRNA	Function unknown	-5.71
AA291377	Hs.50831:23	Homo sapiens Ly-6 antigen/uPA receptor-like domain-containing protein mRNA, complete cds	Function unknown	-6.78

AI420213	Hs.149722:3	cDNA clone IMAGE:2094208 3', mRNA sequence	Function unknown	-8.52
AJ245671	Hs.12844:73	EGFL6, EGF-like-domain; multiple 6	Cell cycle, oncogenesis, integrin ligand, extracellular space. Member of the epidermal growth factor (EGF) repeat superfamily; contains an EGF-like-domain. Expressed early during development, and its expression has been detected in lung and meningioma tumors.	-9.44
AB018305	Hs.5378:149	SPON1, spondin 1, (f-spondin) extracellular matrix protein	Extracellular matrix protein. Very strongly similar to rat F-spondin (Rn.7546); may have a role in the growth and guidance of axons.	-12.55
AW872527	Hs.59761:19	ESTs; Weakly similar to DAP1_HUMAN DEATH-ASSOCIATED PROTEIN 1	Function unknown	-14.17
AF129755	Hs.117772:9, Hs.88474:1	Homo sapiens prostaglandin endoperoxide H synthase-1 mRNA, partial 3' untranslated region.	Function unknown	-21.34
AI023799	Hs.163242:5	Homo sapiens cDNA clone IMAGE:1655725 3' similar to contains MER20.t2 MER20 repetitive element ;, mRNA sequence	Function unknown	-41.34

Table 4 – Mucinous OC

Accession number	UniGene Mapping	Gene symbol and title	Putative Function	P value
AA584890	Hs.5302:132	LGALS4, lectin, galactoside-binding, soluble, 4 (galectin 4)	Lectin, cytosol, cell adhesion, plasma membrane. Binds to beta galactoside, involved in cell adhesion, cell growth regulation, inflammation, immunomodulation, apoptosis and metastasis; member of a family of lectins. bGALS4 is an S-type lectin that is strongly underexpressed in colorectal cancer.	0.0001
AA315933	Hs.120879:17	Homo sapiens, clone MGC:32871 IMAGE:4733535, mRNA, complete cds	Function unknown	0.0001
U47732	Hs.84072:110	TM4SF3, transmembrane 4 superfamily member 3	Integral plasma membrane protein, lysosome, pathogenesis, protein amino acid glycosylation, signal transducer, tumor antigen. Cell surface glycoprotein defined by the monoclonal antibody CO-029 is a 27- to 34-kD membrane protein expressed in gastric, colon, rectal, and pancreatic carcinomas but not in most normal tissues	0.0028
AW503395	Hs.5541:112	ATP2A3, ATPase, Ca++ transporting, ubiquitous	Endoplasmic reticulum, adenosinetriphosphatase, small molecule transport, calcium-transporting ATPase, integral plasma membrane protein. Sarco/endoplasmic reticulum Ca2+-ATPase; pumps calcium.	0.0154
AU076801	Hs.89436:50	CDH17, cadherin 17, LI cadherin (liver-intestine)	Cell adhesion, integral plasma membrane protein, membrane fraction, small molecule transport, transporter. Member of the cadherin family of calcium-dependent glycoproteins; facilitates uptake of peptide-based drugs, may mediate cell-cell interactions. Component of the gastrointestinal tract and pancreatic ducts, intestinal proton-dependent peptide transporter in the first step in oral absorption of many medically important peptide-based drugs.	0.0172
AI073913	Hs.100686:20	LOC155465, anterior gradient protein 3	Oncogenesis	0.0266
AI928445	Hs.92254:80	SYTL2: synaptotagmin-like 2	Synaptotagmin-like protein of the C2 domain-containing family of proteins. Although the specific function of the synaptotagmin-like proteins is unknown, a role in regulation of synaptic vesicle trafficking via their C2 domains has been suggested. Region of weak similarity to murine Gph	0.08
AW088642	Hs.97984:22	SOX17, SRY (sex determining region Y)-box 17	Likely ortholog of mouse SRY-box containing gene 17; alias SOX17	0.1705
W40460	Hs.144442:5	PLA2G10: phospholipase A2, group X	Extracellular, secreted phospholipase A2. Group X secretory phospholipase_a2; hydrolyzes the phospholipid sn-2 ester bond; member of the phospholipase family	0.1888
AA132961	Hs.212533:4	Homo sapiens cDNA: FLJ22572 fis, clone HSI02313	Function unknown	0.1965
AF111856	Hs.105039:48	SLC34A2, solute carrier family 34 (sodium phosphate), member 2	SLC34A2: solute carrier family 34 (sodium phosphate), member 2; contains 8 predicted TMs and a cysteine-rich N-terminal region. Type 2 sodium-dependent phosphate transporter. member of the renal type II co-transporter family.	0.5078

Table 5A – Borderline OC

Accession number	UniGene Mapping	Gene symbol and title	Putative Function	P value
AA143654		zo65a02.r1 Stratagene pancreas (#937208) Homo sapiens cDNA clone IMAGE:591722 5', mRNA sequence	Function unknown	0.036

Table 5B – Borderline OC ratios

Accession number	UniGene Mapping	Gene symbol and title	Putative Function	Ratio
AF196478	Hs.188401:9	ANXA10, annexin A10	Calcium binding. This gene encodes a member of the annexin family. Members of this calcium-dependent phospholipid-binding protein family play a role in the regulation of cellular growth and in signal transduction pathways. The function of this gene has not yet been determined. May have a role in inositol phosphate metabolism.	24.57
AI910275	Hs.1406:49	TFF1, trefoil factor 1 (breast cancer, estrogen-inducible sequence expressed in)	Carbohydrate metabolism, cell growth and/or maintenance, defense response, digestion. Maintains the mucosal surface barrier and stimulates repair processes. Alias trefoil factor, BCE1, human pS2 induced by estrogen from human breast cancer cell line MCF-7. The trefoil factor 1 participates in gastrointestinal cell differentiation by delaying G1-S phase transition and reducing apoptosis (Bossenmeyer-Pourie 2002).	3.90
AW058350	Hs.16762:29	Homo sapiens mRNA; cDNA DKFZp564B2062 (from clone DKFZp564B2062)	function unknown. 98% ID to CLDI_HUMAN CLAUDIN-18	6.56
M29540	Hs.220529:1	CEACAM5 , carcinoembryonic antigen-related cell adhesion molecule 5	Integral plasma membrane protein, tumor antigen. Carcinoembryonic antigen; member of the CEA and immunoglobulin superfamilies. (CEA assay used in monitoring relapse in colon cancer)	5.10
T83911	Hs.11881:66	TM4SF4, transmembrane 4 superfamily member 4	N-linked glycosylation, integral plasma membrane protein, membrane fraction, negative control of cell proliferation. Cell surface glycoprotein:IL-TMP. Member 4 of the transmembrane 4 superfamily (TM4SF); mediates density-associated inhibition of proliferation.	144.32
W95642	Hs.82961:81	TFF3, trefoil factor 3 (intestinal)	Defense response, digestion, extracellular. Trefoil proteins are stable secretory polypeptides that are characterized by the presence of at least 1 copy of a 40-amino acid motif that contains 3 conserved disulfide bonds. May maintain the mucosal surface barrier	3.06

Table 6 – Overall survival

Accession number	UniGene Mapping	Gene symbol	Gene name	Putative Function	P value
AA046217	Hs.105370:2		ESTs	Function unknown	0.00
			NM_015902*:Homo sapiens progestin induced protein (DD5), mRNA. VERSION NM_020967.1 GI	Soluble fraction, cell proliferation, ubiquitin--protein ligase, ubiquitin conjugating enzyme, ubiquitin-dependent protein degradation. Member of the HECT family of proteins; may function as an E3 ubiquitin-protein ligase. This gene is localized to chromosome 8q22, a locus disrupted in a variety of cancers. This gene potentially has a role in regulation of cell proliferation or differentiation.	0.00
T83882	Hs.97927:20		ESTs	Function unknown	0.01
#(NOCAT)			NM_001615*:Homo sapiens actin, gamma 2, smooth muscle, enteric (ACTG2), mRNA. variant 1, mRNA.	Structural protein of muscle. Gamma 2 actin; enteric-type, smooth muscle cell actin.	0.01
AB040888			Homo sapiens mRNA for KIAA1455 protein, partial cds	Function unknown	0.01
AA628980	Hs.192371:3	DSCR8	down syndrome critical region protein DSCR8	Function unknown	0.01
AI623351	Hs.172148:51		ESTs	Function unknown	0.01
AW614420	Hs.204354:383	ARHB	ras homolog gene family, member B	RHO small monomeric GTPase, RHO protein signal transduction, peripheral plasma membrane protein. Ras-related GTP binding protein of the rho subfamily, member B; may regulate assembly of actin stress fibers and focal adhesions; very strongly similar to murine Arhb.	0.01
AA243499	Hs.104800:23		hypothetical protein FLJ10134	Highly similar to murine p19.5; may be a membrane protein	0.01
AF251237	Hs.112208:16	GAGED2	XAGE-1 protein	GAGE genes are expressed in a variety of tumors and in some fetal and reproductive tissues. This gene is strongly expressed in Ewing's sarcoma, alveolar rhabdomyosarcoma and normal testis. The protein encoded by this gene contains a nuclear localization signal and shares a sequence similarity with other GAGE/PAGE proteins. Because of the expression pattern and the sequence similarity, this protein also belongs to a family of CT (cancer-testis) antigens.	0.01
AI970797	Hs.64859:16		ESTs	Function unknown	0.01
AF145713	Hs.61490:51	SCHIP1	schwannomin-interacting protein 1	Cytoplasm. Associates with the neurofibromatosis type 2 protein schwannomin (NF2); contains a coiled-coil domain|Proteome	0.01

X78565	Hs.289114:173 Hs.74637:1	TNC	hexabrachion (tenascin C, cytotactin)	Cell adhesion, extracellular matrix, cell adhesion receptor, ligand binding or carrier. Hexabrachion (tenascin c), an extracellular matrix glycoprotein; has epidermal growth factor-like repeats	0.01
T97307			gb:ye53h05.s1 Soares fetal liver spleen 1NFLS Homo sapiens cDNA clone IMAGE:121497 3', mRNA sequence.	Function unknown	0.01
BE243845	Hs.75511:418	CTGF	connective tissue growth factor	Cell motility, plasma membrane, soluble fraction, response to wounding, extracellular matrix, extracellular space, epidermal differentiation, cell growth and maintenance, insulin-like growth factor binding, insulin-like growth factor receptor binding protein. Connective tissue growth factor; binds IGF, may have a role in regulating normal and neoplastic cell growth	0.01
AW068302	Hs.182183:214 Hs.325474:172 Hs.283080:7	CALD1	caldesmon 1	Cytoskeleton, actin binding, calmodulin binding, tropomyosin binding. Protein of unknown function. Actomyosin regulatory protein, non-muscle form	0.01
AL133561	Hs.241426:5		DKFZP434B061 protein	Function unknown	0.01
BE313555	Hs.7252:158	RAI17	retinoic acid induced 17	Function unknown	0.02
X07820	Hs.2258:1	MMP10	matrix metalloproteinase 10 (MMP10; stromelysin 2)	Zinc binding, extracellular space, extracellular matrix, metalloendopeptidase, proteolysis and peptidolysis. Stromelysin 2; matrix metalloprotease that degrades connective tissue	0.02
AI973016	Hs.15725:77	IER5	immediate early response 5	Function unknown. A related mouse gene may play an important role in mediating the cellular response to mitogenic signals.	0.02
AF084545			Homo sapiens versican Vint isoform, mRNA, partial cds	Function unknown	0.02
U41518	Hs.74602:146 Hs.767:1	AQP1	aquaporin 1 (channel-forming integral protein, 28kD)	Excretion, water transport, water transporter, integral plasma membrane protein. Aquaporin 1 (channel-forming integral protein); member of a family of water-transporters	0.02
Z11894			H.sapiens rearranged mRNA for immunoglobulin kappa chain (VNJ)		0.02
AW138190	Hs.180248:8	ZNF124	zinc finger protein 124 (HZF-16)	DNA binding. C2H2 zinc-finger protein 124	0.02
BE086548	Hs.42346:83 Hs.6975:42	MYOZ2	myozenin 2	calcineurin-binding protein calsarcin-1	0.02
W47196	Hs.166172:50	ARNT	aryl hydrocarbon receptor nuclear translocator	Nucleus, transcription factor, transcription co-activator, transcription, DNA-dependent, protein-nucleus import, translocation, aryl hydrocarbon receptor nuclear translocator. Aryl hydrocarbon receptor nuclear translocator; used in receptor translocation from cytosol to nucleus	0.02
AI796870	Hs.54277:76	DXS9928E	DNA segment on chromosome X (unique) 9928 expressed sequence	Nucleus. Has many charged residues and a possible nuclear localization signal	0.02

X02761	Hs.287820:73 Hs.321592:1	FN1	fibronectin 1	Cell adhesion, cell motility, cell adhesion, soluble fraction, signal transduction, extracellular matrix, extracellular space. Fibronectin 1; member of family of proteins found in plasma and extracellular matrix	0.02
AW968613	Hs.79428:166	BNIP3	BCL2/adenovirus E1B 19kD-interacting protein 3	Anti-apoptosis, apoptosis inhibitor. Bcl2-related protein 3; binds antiapoptotic viral E1B 19 kDa protein and cellular Bcl2 protein	0.02
AW972565	Hs.32399:24		ESTs, Weakly similar to S51797 vasodilator-stimulated phosphoprotein [H.sapiens]	Function unknown	0.02
AF045229	Hs.82280:81	RGS10	regulator of G-protein signalling 10	Regulator of G protein signaling (RGS) family members are regulatory molecules that act as GTPase activating proteins (GAPs) for G alpha subunits of heterotrimeric G proteins. RGS proteins are able to deactivate G protein subunits of the Gi alpha, Go alpha and Gq alpha subtypes. They drive G proteins into their inactive GDP-bound forms.	0.02
AW953853	Hs.292833:19	PAEP	progestagen-associated endometrial protein (placental protein 14, pregnancy-associated endometrial alpha-2-globulin, alpha uterine protein)	Developmental processes. Placental protein 14 (Glycodelin); member of lipocalin superfamily, highly similar to beta-lactoglobulins	0.02
U52426	Hs.74597:75 Hs.157615:3	STIM1	stromal interaction molecule 1	Integral plasma membrane protein, positive control of cell proliferation. Very strongly similar to murine Stim1; may be a transmembrane stromal cell protein	0.02
F08700	Hs.7879:115	IFRD1	interferon-related developmental regulator 1	Myoblast determination. Strongly similar to rat interferon-related developmental regulator 1; may play a role in muscle differentiation	0.02
AI798863	Hs.87191:8		ESTs	Function unknown	0.03
NA			C4001170:gi|6863176|gb|AAF30402.1|AF109924_1 (AF109924) sulfatase 1 precursor [Helix poma		0.03
H52761	Hs.141475:24		Homo sapiens cDNA clone IMAGE:178663	Function unknown	0.03
BE546947	Hs.44276:43	HOXC10	homeo box C10	Embryogenesis and morphogenesis, positive control of cell proliferation, RNA polymerase II transcription factor. Homeobox C10, member of the homeobox developmental regulator family; binds with HOXA13 and HOXC13 to the Lamin B2 origin; ortholog of Drosophila Abdominal-B	0.03
AU076643	Hs.313:257 Hs.329910:1	SPP1	secreted phosphoprotein 1 (osteopontin, bone sialoprotein I, early T-lymphocyte activation 1)	Ossification, extracellular matrix, skeletal development. Osteopontin (bone sialoprotein); bone and blood vessel extracellular matrix protein involved in calcification and atherosclerosis	0.03
#(NOCAT)			NM_015902*:Homo sapiens progestin induced protein (DD5), mRNA. VERSION NM_020967.1 GI		0.03
U20536	Hs.3280:20	CASP6	caspase 6, apoptosis-related cysteine protease	Induction of apoptosis, cysteine-type peptidase, proteolysis and peptidolysis. Caspase 6; a cysteine (thiol) protease; related to the ICE-subfamily of caspases	0.03

AA581602	Hs.41840:7		ESTs	Function unknown	0.03
AJ245210			gb:Homo sapiens mRNA for immunoglobulin gamma heavy chain variable region, partial, clone 1A-4G21.	Function unknown	0.03
X65965			H.sapiens SOD-2 gene for manganese superoxide dismutase		0.03
AI806770	Hs.30258:9		ESTs	Function unknown	0.03
BE386490	Hs.279663:51	PIR	Pirin	Nucleus, transcription co-factor, transcription from Pol II promoter. Putative cofactor of the NFI/CTF1 transcriptional activator	0.03
AW581992	Hs.301434:104 Hs.329017:1	KIAA1387	KIAA1387 protein	Function unknown	0.03
U77534			Human clone 1A11 immunoglobulin variable region (VH5-D-JH4) gene, partial cds	Function unknown	0.03
AL034417	Hs.11169:194 Hs.10958:1 Hs.74137:1		Gene 33/Mig-6	Function unknown	0.03
L10343	Hs.112341:96 Hs.1968:1		Homo sapiens elafin precursor, gene, complete cds	Function unknown	0.03
AW518944	Hs.76325:80 Hs.231299:1	IGJ	immunoglobulin J polypeptide, linker protein for immunoglobulin alpha and mu polypeptides	Linker protein for immunoglobulin alpha and mu polypeptides	0.03
W28729	Hs.236510:6		Human retina cDNA randomly primed sublibrary Homo sapiens cDNA, mRNA sequence	Function unknown	0.03
AI640160	Hs.74131:4	ARSE	arylsulfatase E (chondrodysplasia punctata 1)	Arylsulfatase, skeletal development. Arylsulfatase E; likely involved in warfarin embryopathy.	0.03
U11862	Hs.75741:62	ABP1	amiloride binding protein 1 (amine oxidase (copper-containing))	Metabolism, peroxisome, amine oxidase, drug binding. Diamine oxidase (D-amino-acid oxidase histaminase, amiloride-binding protein); deaminates putrescine and histamine	0.03
AW295980	Hs.252741:3		ESTs	Function unknown	0.03
X59135	Hs.156110:4		H.sapiens mRNA for immunoglobulin 0-81 VL		0.03
BE466173	Hs.379794		Homo sapiens mRNA; cDNA DKFZp686N0118 (from clone DKFZp686N0118)	Function unknown	0.03
#(NOCAT)			Target Exon		0.03
AI354722	Hs.127216:24		hypothetical protein FLJ13465	Function unknown	0.04

M90464	Hs.169825:45 Hs.408:1		Human collagen type IV alpha 5 chain (COL4A5) gene, 5' end	Function unknown	0.04
AA829286	Hs.332053:48 Hs.336462:10	SAA1	serum amyloid A1	Inflammatory response, high-density lipoprotein. Member of the serum amyloid A protein family; member of high density apolipoproteins.	0.04
AI333771	Hs.82204:8 Hs.228363:1		ESTs	Function unknown	0.04
BE465867	Hs.197751:66	DAAM1	dishevelled associated activator of morphogenesis 1	The protein encoded by this gene contains FH domains and belongs to a novel FH protein subfamily implicated in cell polarity, thought to function as a scaffolding protein.	0.04
BE616902	Hs.285313:145 Hs.4055:43	COPEB	core promoter element binding protein	A trancriptional activator, capable of activating transcription approximately 4-fold either on homologous or heterologous promoters. The DNA binding and transcriptional activity of this protein, in conjunction with its expression pattern, suggests that this protein may participate in the regulation and/or maintenance of the basal expression of pregnancy-specific glycoprotein gene and possibly other TATA box-less genes.	0.04
AA430373			gb:zw20f11.s1 Soares ovary tumor NbHOT Homo sapiens cDNA clone IMAGE:769869 3' similar to gb:M63438 IG KAPPA CHAIN PRECURSOR V-III REGION (HUMAN);; mRNA sequence.	Function unknown	0.04
R27430	Hs.271565:3		ESTs	Function unknown	0.04
BE387335	Hs.283713:68	CTHRC1	collagen triple helix repeat containing 1	Function unknown	0.04
AW264102	Hs.39168:16		ESTs	Function unknown	0.04
NA			Target Exon	Function unknown	0.04
AW952323	Hs.129908:39		KIAA0591 protein	Function unknown	0.04
AA088177	Hs.172870:13		ESTs	Function unknown	0.04
BE614567	Hs.19574:123	MGC5469	hypothetical protein MGC5469	Function unknown	0.04
AL079658	Hs.338207:139 Hs.146559:1	FRAP1	FK506 binding protein 12-rapamycin associated protein 1	DNA repair, DNA recombination, cell cycle control, 1-phosphatidylinositol 3-kinase, inositol/phosphatidylinositol kinase. FKBP-rapamycin associated protein; phosphatidylinositol kinase that may mediate rapamycin inhibition of the cell cycle progression through G1	0.04
NM_002776	Hs.69423:46 Hs.275464:1	KLK10	kallikrein 10 (KLK10) (PRSSL1) (nes1)	Extracellular, serine-type peptidase. Putative serine protease	0.04
BE261944	Hs.118625:62	CYB561	cytochrome b-561	Energy pathways, secretory vesicle, cytochrome b5 reductase, secretory vesicle membrane, integral plasma membrane protein. Cytochrome b561; serves as a biological marker for adrenergic secretory vesicles	0.04

NM_006379	Hs.171921:50	SEMA3C	sema domain, immunoglobulin domain (Ig), short basic domain, secreted, (semaphorin) 3C	Drug resistance, immune response, cell growth and maintenance. Semaphorin E; member of a protein family involved in neuronal growth cone guidance	0.04
AI002238	Hs.11482:19	SFRS11	splicing factor, arginine/serine-rich 11	Nucleus, mRNA splicing, mRNA processing, pre-mRNA splicing factor, May have a role in pre-mRNA splicing; contains arginine/serine-rich domain and an RRM domain	0.04
#(NOCAT)			ENSP00000231844*:Ecotropic virus integration 1 site protein.		0.04
X81789	Hs.77897:149	SF3A3	splicing factor 3a, subunit 3, 60kD	Nucleus, spliceosome, mRNA splicing, mRNA processing, pre-mRNA splicing factor. Spliceosome-associated protein 3a, subunit 3; component of the essential heterotrimeric splicing factor SF3a; contains a zinc finger	0.04
AA381553	Hs.198253:21	HLA-DQA1	major histocompatibility complex, class II, DQ alpha 1	Pathogenesis, class II major histocompatibility complex antigen. Alpha 1 chain of HLA-DQ1 class II molecule (Ia antigen); complex binds peptides and presents them to CD4+ T lymphocytes|Proteome	0.04
AB001914			Homo sapiens PACE4 gene, exon 23-25, complete cds	Function unknown	0.04
AA311919	Hs.69851:24	NOLA1	nucleolar protein family A, member 1 (H/ACA small nucleolar RNPs)	Involved in various aspects of rRNA processing and modification. Localize to the dense fibrillar components of nucleoli and to coiled (Cajal) bodies in the nucleus.	0.04
AI381750	Hs.283437:122 Hs.10065:58		HTGN29 protein	Function unknown	0.04
#(NOCAT)			NM_000636*:Homo sapiens superoxide dismutase 2, mitochondrial (SOD2), mRNA. expression) (RFX2), mRNA.	Mitochondrion, oxidative stress response, manganese superoxide dismutase. Manganese superoxide dismutase; intramitochondrial free radical scavenging enzyme; has strong similarity to murine Sod2.	0.04
AA292998	Hs.163900:25		ESTs	Function unknown	0.04
BE439580	Hs.75498:40	SCYA20	small inducible cytokine subfamily A (Cys-Cys), member 20	Chemokine, chemotaxis, immune response, signal transduction, extracellular space, cell-cell signalling, inflammatory response, antimicrobial humoral response. Cytokine A20 (exodus); chemotactic factor for lymphocytes, but not a chemotactic factor for monocytes	0.04
AI677897	Hs.76640:124	RGC32	RGC32 protein	Cytoplasm, cell cycle regulator, regulation of CDK activity. Strongly similar to RGC-32.	0.04
#(NOCAT)			Target Exon	Function unknown	0.04
N72403			Homo sapiens cDNA clone IMAGE:245132	Function unknown	0.05
BE003054	Hs.1695:46	MMP12	matrix metalloproteinase 12 (macrophage elastase)	Zinc binding, cell motility, macrophage elastase, extracellular matrix, proteolysis and peptidolysis. Matrix metalloprotease; degrades elastin	0.05

AL035588	Hs.153203:26 Hs.23391:1		Human DNA sequence from clone 696P19 on chromosome 6p12.3-21.2. Contains the gene for TFEB, an NPM1 (Nucleophosmin, Numatrin) pseudogene and the MDFI gene for MyoD family inhibitor (myogenic repressor I-MF). Contains ESTs, STSs, GSSs and two putative CpG islands, complete sequence	Function unknown	0.05
AI080491	Hs.93270:3		ESTs, Moderately similar to S65657 alpha-1C-adrenergic receptor splice form 2 [H.sapiens]	Function unknown	0.05
AW770994	Hs.30340:125		hypothetical protein KIAA1165	Function unknown	0.05
H24177	Hs.75262:69 Hs.238912:1	CTSO	cathepsin O	Cysteine-type endopeptidase, proteolysis and peptidolysis. Cathepsin O; cysteine (thiol) protease	0.05
AF146761	Hs.20450:29		BCM-like membrane protein precursor	Function unknown	0.05
NM_001955	Hs.2271:45 Hs.306:1	EDN1	endothelin 1	Circulation, peptide hormone, soluble fraction, signal transduction, extracellular space, cell-cell signalling, blood pressure regulation, positive control of cell proliferation. Preproendothelin 1; precursor of the hormone endothelin 1	0.05
AI680737	Hs.289068:204 Hs.326198:1	TCF4	transcription factor 4	Nucleus, RNA polymerase II transcription factor, transcription regulation from Pol II promoter. Transcriptional activator; interacts with ITF1 (TCF3); contains basic helix-loop-helix domain|Proteome	0.05
AI752666	Hs.76669:183	NNMT	nicotinamide N-methyltransferase	Nicotinamide N-methyltransferase; catalyzes the N-methylation of nicotinamide and other pyridines, structurally-related drugs and xenobiotics|Proteome	0.05
AA505445	Hs.300697:21	IGHG3	immunoglobulin heavy constant gamma 3 (G3m marker)	Constant region of heavy chain of IgG3	0.05

Table 7 – Clinical recurrence

Accession number	UniGene Mapping	Gene symbol	Gene name	Putative Function	P value
M86849	Hs.323733:62, Hs.300816:5	GJB2	gap junction protein, beta 2, 26kD (connexin 26)	Hearing, connexon, plasma membrane, connexon channel, cell-cell signalling, small molecule transport. Connexin 26; gap junction protein expressed in various tissues including cochlea.	0.00
AW963419	Hs.155223:20	STC2	stanniocalcin 2	Peptide hormone, cell-cell signalling, glycopeptide hormone, nutritional response pathway, cell surface receptor linked signal transduction. Stanniocalcin 2; may regulate metal ion homeostasis and inhibits phosphate uptake.	0.00
BE298665	Hs.14846:132		Homo sapiens mRNA; cDNA DKFZp564D016 (from clone	Function unknown	0.00
AK000637	Hs.46624:11	HSPC043	HSPC043 protein	Function unknown	0.00
BE077546	Hs.31447:27		ESTs, Moderately similar to A46010 X-linked retinopathy protein [H.sapiens]	Function unknown	0.00
T97307			gb:ye53h05.s1 Soares fetal liver spleen 1NFLS Homo sapiens cDNA clone IMAGE:121497 3', mRNA sequence.	Function unknown	0.00
R24601	Hs.108300:46		Homo sapiens adenylosuccinate synthetase isozyme (ADSS) mRNA, complete cds	Function unknown	0.00
BE090176	Hs.179902:95		Interim-CDw92 antigen	choline transporter-like protein	0.00
AA393907	Hs.97179:22		ESTs	Function unknown	0.00
W28729	Hs.236510:6		Homo sapiens mRNA; cDNA DKFZp666D074 (from clone DKFZp666D074)	Function unknown	0.00
BE313754	Hs.13350:52		Homo sapiens mRNA; cDNA DKFZp586D0918	Function unknown	0.01
AW673081	Hs.54828:9		ESTs	Function unknown	0.01
AA356694	Hs.94011:42, Hs.7744:2, Hs.231043:1	HCA4	Hepatocellular carcinoma-associated protein HCA4	Function unknown	0.01
L08239	Hs.5326:11	MG61	porcupine	amino acid system N transporter 2;	0.01
BE397649	Hs.94109:40		Homo sapiens cDNA FLJ34399 fis, clone HCHON2001359	Function unknown	0.01
NM_012317	Hs.45231:36	LDOC1	leucine zipper, down-regulated in cancer 1	Nucleus, negative control of cell proliferation. Nuclear protein; contains a leucine zipper-like motif	0.01

NM_000947	Hs.74519:20	PRIM2A	primase, polypeptide 2A (58kD)	DNA primase, DNA replication, priming, alpha DNA polymerase:primase complex. Subunit of DNA primase polypeptide 2A; part of the DNA polymerase alpha-primase complex	0.01
AJ250562	Hs.82749:133		Homo sapiens partial TM4SF2 gene for tetraspanin protein, exon 1 and joined CDS	Function unknown	0.01
AL040183	Hs.123484:24, Hs.326906:1		Homo sapiens mRNA; cDNA DKFZp686E1934 (from clone DKFZp686E1934)	Function unknown	0.01
BE207573	Hs.83321:32	NMB	neuromedin B	Peptide hormone, soluble fraction, signal transduction, cell-cell signalling. Precursor of neuromedin B, a C-terminally amidated peptide hormone; similar to bombesin	0.01
BE564162	Hs.250820:45	FLJ14827	hypothetical protein FLJ14827	Function unknown	0.01
BE439580	Hs.75498:40	SCYA20	small inducible cytokine subfamily A (Cys-Cys), member 20	Chemokine, chemotaxis, immune response, signal transduction, extracellular space, cell-cell signalling, inflammatory response, antimicrobial humoral response. Cytokine A20 (exodus); chemotactic factor for lymphocytes, but not a chemotactic factor for monocytes	0.01
AW067800	Hs.155223:52	STC2	stanniocalcin 2	Peptide hormone, cell-cell signalling, glycopeptide hormone, nutritional response pathway, cell surface receptor linked signal transduction. Stanniocalcin 2; may regulate metal ion homeostasis and inhibits phosphate uptake.	0.01
AA569756	Hs.87803:10		Homo sapiens cDNA FLJ30156 fis, clone BRACE2000487	Function unknown	0.01
AW138190	Hs.180248:8	ZNF124	zinc finger protein 124 (HZF-16)	DNA binding. C2H2 zinc-finger protein 124	0.01
AF126245	Hs.14791:48	ACAD8	acyl-Coenzyme A dehydrogenase family, member 8	Lipid metabolism, acyl-CoA dehydrogenase. Member of the acyl-Coenzyme A dehydrogenase family; alpha,beta-dehydrogenates acyl-CoA esters	0.01
L10343	Hs.112341:96, Hs.1968:1		Homo sapiens elafin precursor, gene, complete cds	elastase-specific inhibitor in bronchial secretions	0.01
NM_002514	Hs.235935:38	NOV	nephroblastoma overexpressed gene	Insulin-like growth factor receptor binding protein. Insulin-like growth factor binding protein; may play a role in nephrogenesis	0.01
AI863735	Hs.186755:3		ESTs	Function unknown	0.01
NM_005397	Hs.16426:160, Hs.248780:1	PODXL	podocalyxin-like	Integral plasma membrane protein. Transmembrane protein similar to rodent podocalyxins	0.01
W26391	Hs.301206:100	KIF3B	kinesin family member 3B	Plus-end kinesin, microtubule motor, anterograde axon cargo transport, plus-end-directed kinesin ATPase, determination of left-right asymmetry. Similar to murine Kif3b; may have a role in intracellular organelle transport, may act in left-right determination in embryogenesis; may be a microtubule-associated motor protein	0.01
H15474	Hs.132898:156	FADS1	fatty acid desaturase 1	C-5 sterol desaturase, fatty acid desaturation, integral membrane protein. Delta-5 desaturase; catalyzes production of polyenoic fatty acids such as arachidonic acid	0.01

U51166	Hs.173824:106	TDG	thymine-DNA glycosylase	DNA repair, nucleoplasm, damaged DNA binding, base-excision repair, G/T-mismatch-specific thymine-DNA glycosylase. Thymine-DNA glycosylase; excises uracil and thymine from mispairs with guanidine	0.01
AA243499	Hs.104800:23	FLJ10134	hypothetical protein FLJ10134	Highly similar to murine p19.5; may be a membrane protein	0.01
AW408807	Hs.34497:46	FLJ22116	hypothetical protein FLJ22116	Function unknown	0.01
AI738719	Hs.198427:98	HK2	hexokinase 2	Hexokinase, cell cycle control, glucose catabolism, glucose metabolism, mitochondrial outer membrane. Hexokinase II; converts aldo- and keto-hexose sugars to the hexose-6-phosphate	0.01
AB040888	Hs.41793:110		Homo sapiens mRNA for KIAA1455 protein, partial cds	Function unknown	0.01
BE313077	Hs.93135:40, Hs.228357:1		Homo sapiens cDNA FLJ39971 fis, clone SPLEN2028066	Function unknown	0.01
AI677897	Hs.76640:124	RGC32	RGC32 protein	Cytoplasm, cell cycle regulator, regulation of CDK activity. Strongly similar to RGC-32	0.01
C14898	Hs.192986:5		ESTs	Function unknown	0.01
AI821730	Hs.116524:7		Homo sapiens cDNA FLJ35800 fis, clone TESTI2005933	Function unknown	0.01
AF007393	Hs.177574:111	PRKRIR	protein-kinase, interferon-inducible double stranded RNA dependent inhibitor, repressor of (P58 repressor)	Stress response, protein binding, signal transduction, translational regulation, negative control of cell proliferation. Regulates interferon-induced protein kinase PKR (PRKR) activity by binding and inhibiting the PKR-regulator P58IPK (PRKRI)	0.01
H65423	Hs.17631:42	DKFZP434E2135	hypothetical protein DKFZp434E2135	Function unknown	0.01
N46243	Hs.110373:26		ESTs, Highly similar to T42626 secreted leucine-rich repeat-containing protein SLIT2 - mouse (fragment) [M.musculus]	Function unknown	0.01
AA095971	Hs.198793:56, Hs.309674:7		Homo sapiens cDNA: FLJ22463 fis, clone HRC10126	Function unknown	0.01
U20350	Hs.78913:33	CX3CR1	chemokine (C-X3-C) receptor 1	Virulence, chemotaxis, coreceptor, cell adhesion, plasma membrane, chemokine receptor, response to wounding, cellular defense response, integral plasma membrane protein, G-protein linked receptor protein signalling pathway. CX3C chemokine receptor; G protein-coupled receptor, mediates leukocyte migration and adhesion, binds the CX3C chemokine fractalkine and signals through a pertussis toxin sensitive G-protein	0.01
NM_005756	Hs.184942:18	GPR64	G protein-coupled receptor 64	Spermatogenesis, G-protein linked receptor, integral plasma membrane protein, G-protein linked receptor protein signalling pathway. Member of the G protein-coupled receptor family	0.01
D19589	Hs.13453:87	FLJ14753	hypothetical protein FLJ14753	Function unknown	0.02

AW957446	Hs.301711:74		ESTs	Function unknown	0.02
AW294647	Hs.233634:40	C20orf39	chromosome 20 open reading frame 39	Function unknown	0.02
BE159718	Hs.85335:46		Homo sapiens, clone IMAGE:4513159, mRNA	Function unknown	0.02
AI888490	Hs.55902:22	EDG3	endothelial differentiation, sphingolipid G-protein-coupled receptor, 3	Lipid binding, plasma membrane, inflammatory response, G-protein linked receptor, embryogenesis and morphogenesis, integral plasma membrane protein, positive control of cell proliferation, cytostolic calcium ion concentration elevation, G-protein linked receptor protein signalling pathway. Lysosphingolipid receptor, a G protein-coupled receptor; activates calcium flux and serum response element driven transcription	0.02
AA022569	Hs.29802:35, Hs.271785:1		ESTs	Function unknown	0.02
BE147740	Hs.104558:21		ESTs, Moderately similar to hypothetical protein FLJ20378 [Homo sapiens]	Function unknown	0.02
AI798863	Hs.87191:8		ESTs	Function unknown	0.02
BE464341	Hs.21201:18		Interim-DKFZP566B0846: nectin 3	Low similarity to PVRL1; may be a membrane glycoprotein; contains an immunoglobulin (Ig) domain	0.02
AL080235	Hs.35861:34, Hs.289068:1	RIS1	Ras-induced senescence 1	Rat brain specific binding protein	0.02
AI557212	Hs.17132:102, Hs.330782:1		ESTs	Function unknown	0.02
X75208	Hs.2913:41	EPHB3	EphB3	Signal transduction, integral plasma membrane protein, transmembrane receptor protein tyrosine kinase. Eph-related receptor tyrosine kinase B3	0.02
AA628980	Hs.192371:3	DSCR8	down syndrome critical region protein DSCR8	Melanoma-testis-associated protein 2	0.02
BE242587	Hs.118651:39	HHEX	hematopoietically expressed homeobox	Nucleus, DNA binding, transcription factor, developmental processes, antimicrobial humoral response. Member of the homeodomain family of DNA binding proteins; may regulate gene expression, morphogenesis, and differentiation	0.02
NM_005512	Hs.151641:65	GARP	glycoprotein A repetitions predominant	Integral plasma membrane protein. Putative transmembrane cell surface protein; has an extracellular domain comprised largely of leucine-rich repeats	0.02
AW953853	Hs.292833:19	PAEP	progestagen-associated endometrial protein (placental protein 14, pregnancy-associated endometrial alpha-2-globulin, alpha uterine protein)	Developmental processes. Placental protein 14 (Glycodelin); member of lipocalin superfamily, highly similar to beta-lactoglobulins	0.02

AU076611	Hs.154672:122	MTHFD2	methylene tetrahydrofolate dehydrogenase (NAD dependent), methenyltetrahydrofolate cyclohydrolase	Mitochondrion, electron transporter, methenyltetrahydrofolate cyclohydrolase, methylenetetrahydrofolate dehydrogenase. NAD-dependent methylene tetrahydrofolate dehydrogenase-cyclohydrolase; may provide formyltetrahydrofolate for formylmethionyl tRNA synthesis; involved in initiation of mitochondrial protein synthesis	0.02
AW968613	Hs.79428:166	BNIP3	BCL2/adenovirus E1B 19kD-interacting protein 3	Anti-apoptosis, apoptosis inhibitor. Bcl2-related protein 3; binds antiapoptotic viral E1B 19 kDa protein and cellular Bcl2 protein	0.02
AL353944	Hs.50115:14		Homo sapiens mRNA; cDNA DKFZp761J1112 (from clone DKFZp761J1112)	Function unknown	0.02
BE614149	Hs.20814:29, Hs.306626:27		LOC51072: C21orf19-like protein	Function unknown	0.02
AA292998	Hs.163900:25		ESTs	Highly similar to winged helix/forkhead transcription factor	0.02
H12912	Hs.274691:138	AK3	adenylate kinase 3	Nucleobase, nucleoside, nucleotide and nucleic acid metabolism. Adenylate kinase 3; strongly similar to murine Ak4	0.02
AA188763	Hs.36793:4	SLC12A8	solute carrier family 12 (potassium/chloride transporters), member 8	Solute carrier family 12 (potassium/chloride transporters), member 8	0.02
AK000596	Hs.3618:56	HPCAL1	hippocalcin-like 1	Calcium-binding protein with similarity to hippocalin (human HPCA); expressed only in the brain.	0.02
AI970797	Hs.64859:16		ESTs	Function unknown	0.02
AW519204	Hs.40808:22		ESTs	Function unknown	0.02
Z42387	Hs.83883:114	TMEPAI	transmembrane, prostate androgen induced RNA	Function unknown	0.02
AF145713	Hs.61490:51	SCHIP1	schwannomin-interacting protein 1	Cytoplasm. Associates with the neurofibromatosis type 2 protein schwannomin (NF2); contains a coiled-coil domain	0.02
AA972412	Hs.13755:41	FBXW2	f-box and WD-40 domain protein 2	Protein modification, ubiquitin--protein ligase, proteolysis and peptidolysis, ubiquitin conjugating enzyme. F-box and WD-40 domain protein 2; putative SCF ubiquitin ligase subunit involved in protein degradation; contains a WD-40 domain and an F-box	0.02
AK001564	Hs.104222:139, Hs.296267:4		Homo sapiens cDNA FLJ10702 fis, clone NT2RP3000759, weakly similar to ADP-RIBOSYLATION FACTOR	Member of the ADP-ribosylation factor (ARF) family; putative GTP-binding protein involved in protein trafficking	0.02
AW959861	Hs.290943:28		ESTs	Function unknown	0.02
BE313555	Hs.7252:158	RAI17	retinoic acid induced 17	Function unknown	0.02
W25005	Hs.24395:199		zb67e02.r1 Soares_fetal_lung_NbHL19W Homo sapiens cDNA clone IMAGE:308666 5', mRNA sequence	Function unknown	0.02
AI193356	Hs.160316:3		ESTs	Function unknown	0.02

AF111106	Hs.3382:223	PPP4R1	protein phosphatase 4, regulatory subunit 1	Protein phosphatase	0.02
AI130740	Hs.6241:116	PIK3R1	phosphoinositide-3-kinase, regulatory subunit, polypeptide 1 (p85 alpha)	A family of enzymes that phosphorylate the 3'-hydroxyl of phosphatidylinositol (Ptdins).	0.02
AA985190	Hs.246875:42	FLJ20059	hypothetical protein FLJ20059	Contains four Kelch motif domains	0.02
BE221880	Hs.268555:144	XRN2	5'-3' exoribonuclease 2	Nucleus, nuclease, recombination, RNA catabolism, RNA processing. 5'-3' Exoribonuclease; similar to Schizosaccharomyces pombe Dhp1p	0.03
AF084545			Homo sapiens versican Vint isoform, mRNA, partial cds	Function unknown	0.03
R26584	Hs.267993:43		TAPBP-R: TAP binding protein related	Has low similarity to TAPBP (Tapasin); contains two immunoglobulin (Ig) domains|Proteome	0.03
AW247380	Hs.12124:116	ELAC2	elaC homolog 2 (E. coli)	putative prostate cancer susceptibility protein	0.03
AA364261	Hs.131365:7		ESTs	Weakly similar to T31613 hypothetical protein Y50E8A.i - Caenorhabditis elegans [C.elegans]	0.03
U25849	Hs.75393:141	ACP1	Human red cell-type low molecular weight acid phosphatase (ACP1) gene, exon 6 and 7, complete cds	Acid phosphatase	0.03
AF262992	Hs.123159:14	SPAG4	sperm associated antigen 4	Spermatogenesis, structural protein. Sperm associated antigen 4; predicted ortholog of rat SPAG4, which interacts with rat ODF27, the 27kDa outer dense fiber protein of elongating spermatids	0.03
AW342140	Hs.182545:1		ESTs, Weakly similar to POL2_MOUSE Retrovirus-related POL polyprotein	Function unknown	0.03
AL133572	Hs.199009:58	PCCX2	protein containing CXXC domain 2	DNA-binding protein with PHD finger and CXXC domain, is regulated by proteolysis.	0.03
AI497778	Hs.20509:4	HBXAP	hepatitis B virus x associated protein	Weakly similar to Drosophila CG8677	0.03
AI745379	Hs.42911:31	TAF13	TAF13 RNA polymerase II, TATA box binding protein (TBP)-associated factor, 18 kD	TFIID complex, protein binding, transcription factor, general RNA polymerase II transcription factor. TBP-associated factor K; component of TFIID complexes containing TAFII30 (TAF2H)	0.03
U51712	Hs.13775:135		LAGY: lung cancer-associated Y protein	The protein encoded by this gene is a lung cancer associated protein. The function of the protein is not known. Multiple alternatively spliced transcript variants have been described for this gene but some of their full length sequence has not been determined.	0.03
AW375974	Hs.156704:4		ESTs	Function unknown	0.03

AF251237	Hs.112208:16	GAGED2	G antigen, family D, 2	GAGE genes are expressed in a variety of tumors and in some fetal and reproductive tissues. This gene is strongly expressed in Ewing's sarcoma, alveolar rhabdomyosarcoma and normal testis. The protein encoded by this gene contains a nuclear localization signal and shares a sequence similarity with other GAGE/PAGE proteins. Because of the expression pattern and the sequence similarity, this protein also belongs to a family of CT (cancer-testis) antigens.	0.03

Table 8 – Biochemical recurrence

Accession number	UniGene Mapping	Gene symbol	Gene name	Putative Function	P value
			NM_000636*:Homo sapiens superoxide dismutase 2, mitochondrial (SOD2), mRNA. expression) (RFX2), mRNA.	Mitochondrion, oxidative stress response, manganese superoxide dismutase. Manganese superoxide dismutase; intramitochondrial free radical scavenging enzyme; has strong similarity to murine Sod2.	0.02
AA130986	Hs.271627:1		ESTs	Function unknown	0.01
AA216363	Hs.262958:48, Hs.327737:2	DKFZP434B044	hypothetical protein DKFZp434B044	Function unknown	0.01
AA628980	Hs.192371:3	DSCR8	down syndrome critical region protein DSCR8	Function unknown	0.00
AA811657	Hs.220913:9		Homo sapiens cDNA FLJ40827 fis, clone TRACH2011500	Function unknown	0.02
AA897108			gb:am08a06.s1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone 3', mRNA sequence	Function unknown	0.01
AB040888	Hs.41793:110		Homo sapiens mRNA for KIAA1455 protein, partial cds	Function unknown	0.02
AF212225	Hs.283693:104		Homo sapiens BM022 mRNA, complete cds	Function unknown	0.02
AI089575	Hs.9071:52		ESTs	Function unknown	0.02
AI282028	Hs.25205:10		ESTs	Function unknown	0.02
AI368826	Hs.30654:15		FLJ10849: hypothetical protein FLJ10849	Moderately similar to members of the septin family	0.02
AI718702	Hs.308026:11, Hs.194490:6	HLA-DRB3	major histocompatibility complex, class II, DR beta 5	Signal transduction, integral plasma membrane protein, class II major histocompatibility complex antigen. Beta 3 chain of HLA-DR; subunit of MHC class II molecule, complex binds peptides and presents them to CD4+ T lymphocytes	0.02
AI827248	Hs.224398:3		Homo sapiens cDNA FLJ11469 fis, clone HEMBA1001658	Function unknown	0.01
AK002039	Hs.26243:38	MRVI1	murine retrovirus integration site 1 homolog	Oncogenesis, tumor suppressor, endoplasmic reticulum membrane. Similar to human MLRP; may act as a tumor suppressor	0.02
AL109791	Hs.241559:3		Homo sapiens mRNA full length insert cDNA clone EUROIMAGE 151432	Function unknown	0.00
AW090198	Hs.4779:29		LOC127829: hypothetical protein BC015408	Function unknown	0.01
AW296454	Hs.24743:92		FLJ20171: hypothetical protein FLJ2017	Contains three RNA recognition motifs (RRM, RBD, or RNP)	0.02

AW445034	Hs.256578:4		ESTs	Function unknown	0.00
AW452948	Hs.257631:3		ESTs	Function unknown	0.01
AW470411	Hs.288433:27		HNT: neurotrimin	Cell adhesion, neuronal cell recognition, integral plasma membrane protein. Neurotrimin; may function as a GPI-anchored neural cell adhesion molecule; member of the immunoglobulin superfamily	0.02
AW885727	Hs.301570:22	FST	follistatin	Developmental processes. Follistatin; inhibits the release of follicle-stimulating hormone (FSH)	0.01
AW970859	Hs.313503:4		ESTs	Function unknown	0.02
AW979189	Hs.283367:3		ESTs	Function unknown	0.01
BE165866	Hs.83623:66		Human XIST, coding sequence "a" mRNA (locus DXS399E)	XIST mRNA	0.01
BE175582			gb:RC5-HT0580-100500-022-C01 HT0580 Homo sapiens cDNA, mRNA sequence	Function unknown	0.01
BE242587	Hs.118651:39	HHEX	hematopoietically expressed homeobox	Nucleus, DNA binding, transcription factor, developmental processes, antimicrobial humoral response. Member of the homeodomain family of DNA binding proteins; may regulate gene expression, morphogenesis, and differentiation|	0.01
BE271927	Hs.87385:31, Hs.307940:4		LOC115416: hypothetical protein BC012331	Function unknown	0.01
BE439580	Hs.75498:40	SCYA20	small inducible cytokine subfamily A (Cys-Cys), member 20	Chemokine, chemotaxis, immune response, signal transduction, extracellular space, cell-cell signalling, inflammatory response, antimicrobial humoral response. Cytokine A20 (exodus); chemotactic factor for lymphocytes, but not a chemotactic factor for monocytes	0.02
BE464016	Hs.238956:35		Homo sapiens cDNA FLJ37793 fis, clone BRHIP3000473	Function unknown	0.02
D63216	Hs.153684:137	FRZB	frizzled-related protein	Membrane, extracellular, skeletal development. Frizzled-related protein; similar to frizzled family of receptors	0.02
F34856	Hs.292457:120		Homo sapiens, clone MGC:16362 IMAGE:3927795, mRNA, complete cds	Function unknown	0.02
M83822	Hs.62354:112	LRBA	LPS-responsive vesicle trafficking, beach and anchor containing	May mediate protein-protein interactions; contains two WD domains (WD-40 repeats) and a beige/BEACH domain|Proteome	0.02
N33937	Hs.10336:6		ESTs	Function unknown	0.01
N49068	Hs.93966:4		ESTs	Function unknown	0.01
N51357	Hs.260855:62		NSE1: NSE1	Function unknown	0.02
N80486	Hs.39911:17		Homo sapiens mRNA for FLJ00089 protein,	Function unknown	0.02

			partial cds		
NM_000954	Hs.8272:265, Hs.332355:1	PTGDS	prostaglandin D2 synthase (21kD, brain)	Membrane, prostaglandin-D synthase. Glutathione-independent prostaglandin D2 synthase; membrane associated, catalyzes synthesis of prostaglandin D; member of the lipocalin family of transporters	0.02
NM_005756	Hs.184942:18	GPR64	G protein-coupled receptor 64	Spermatogenesis, G-protein linked receptor, integral plasma membrane protein, G-protein linked receptor protein signalling pathway. Member of the G protein-coupled receptor family	0.02
NM_016652	Hs.268281:61	CRNKL1	Crn, crooked neck-like 1 (Drosophila)	Function unknown	0.02
R26584	Hs.267993:43		TAPBP-R: TAP binding protein related	Has low similarity to TAPBP (Tapasin); contains two immunoglobulin (Ig) domains	0.01
R31178	Hs.287820:6	FN1	fibronectin 1	Cell adhesion, cell motility, cell adhesion, soluble fraction, signal transduction, extracellular matrix, extracellular space. Fibronectin 1; member of family of proteins found in plasma and extracellular matrix	0.02
W05391	Hs.83623:8		Homo sapiens cDNA FLJ30298 fis, clone BRACE2003172	Function unknown	0.02
W25005	Hs.24395:199		zb67e02.r1 Soares_fetal_lung_NbHL19W Homo sapiens cDNA clone IMAGE:308666 5', mRNA sequence	Function unknown	0.01
W45393	Hs.55888:15	ATF7	activating transcription factor 7	Transcription factor. Leucine zipper DNA-binding protein; recognizes a cAMP response element (CRE), involved in the regulation of adenovirus Ela-responsive and cellular cAMP-inducible promoters	0.02
W68815	Hs.301885:20		Homo sapiens cDNA FLJ33794 fis, clone CTONG1000009	Function unknown	0.01
X65965			H.sapiens SOD-2 gene for manganese superoxide dismutase	Mitochondrion, oxidative stress response, manganese superoxide dismutase. Manganese superoxide dismutase; intramitochondrial free radical scavenging enzyme; has strong similarity to murine Sod2.	0.01
X76732	Hs.3164:58	NUCB2	nucleobindin 2	Cytosol, DNA binding, plasma membrane, calcium binding, extracellular space. Nucleobindin 2; may bind DNA and calcium; has DNA-binding and EF-hand domains, and a leucine-zipper	0.02
Z45051	Hs.22920:25	C20orf103	chromosome 20 open reading frame 103	Low similarity to a region of murine Lamp1|Proteome	0.02